

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 16:58:16 ; Search time 626 Seconds
(without alignments)
181.293 Million cell updates/sec

Title: US-09-831-683B-10

Perfect score: 3734
Sequence: 1 MEALSPKSDPPISIIICA.....SSKEPVLPALPDGRQASLS 715

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3734	100.0	715	21	AA194276
2	2657	71.2	716	21	AA194276
3	2657	71.2	719	21	AA194276
4	2657	71.2	748	21	AA194276
5	1537	41.2	725	23	ABP73689
6	1529	40.3	1714	22	ABP73689
7	1506	40.3	1512	22	AAU04349
8	1502	40.2	1512	22	AAU04349
9	1493	40.0	1550	22	AAU04349

10	1471.5	39.4	715	24	ABJ25717
11	1465.5	39.2	715	24	ABJ25717
12	697.5	18.7	581	22	AAU0163
13	666.5	17.8	562	24	ABU06096
14	661.5	17.7	573	24	ABP77077
15	653	17.5	577	22	AA196155
16	599	16.0	786	23	ABP99409
17	588	15.7	799	23	ABP73620
18	561.5	15.0	775	21	AA192292
19	553	14.8	759	22	AB195073
20	280	7.5	156	22	AAU23511
21	255.5	6.8	484	24	ABP80994
22	249.5	6.7	511	22	AB196070
23	243	6.5	463	19	AA198559
24	243	6.5	463	22	AAU35732
25	232	6.2	480	22	AAU35406
26	230.5	6.2	481	23	ABP28650
27	229	6.1	491	22	AAU35347
28	227.5	6.1	471	22	AAU34644
29	225	6.0	482	22	AAU33457
30	225	6.0	483	23	AB195407
31	220.5	5.9	491	23	AB194154
32	216.5	5.8	484	23	ABP30150
33	215.5	5.8	503	23	ABP28649
34	215	5.8	404	21	AA194278
35	213.5	5.7	486	24	AAU02579
36	212.5	5.7	516	20	AAU35180
37	212	5.7	80	23	ABP02943
38	212	5.7	481	22	AAU34088
39	212	5.7	484	18	AA192187
40	212	5.7	487	22	AAU36863
41	207.5	5.6	478	23	ABP30889
42	206	5.5	486	22	AAU37862
43	205.5	5.5	486	22	AAU38363
44	205.5	5.5	480	20	AAU09552
45	205.5	5.5	480	21	AAU32829

ALIGNMENTS

RESULT 1	AA194276	standard; Protein: 715 AA.
ID	AA194276	
XX	AA194276	
AC	AA194276	
XX	AA194276	
DT	10-AUG-2000	(first entry)
XX	10-AUG-2000	
DE	Corn glutamyl-tRNA synthetase.	
XX	Corn; glutamyl-tRNA synthetase; aminoacyl-tRNA synthetase;	
KW	AARS; herbicide; plant toxin; protein synthesis inhibition; enzyme.	
KW	AARS; herbicide; plant toxin; protein synthesis inhibition; enzyme.	
XX	AARS; herbicide; plant toxin; protein synthesis inhibition; enzyme.	
OS	Zea mays.	
XX	Zea mays.	
FN	W0200028057-A2.	
XX	W0200028057-A2.	
PD	18-MAY-2000.	
XX	18-MAY-2000.	
PF	09-NOV-1999;	99NO-US26478.
XX	09-NOV-1999;	
FR	10-NOV-1998;	98US-0107789.
XX	10-NOV-1998;	
PA	(DUPO) DU PONT DE NEMOURS & CO E I.	
PA	(PION-) PIONEER HI-BRED INT INC.	
PI	Famodu CO, Simmons C;	
XX	Famodu CO, Simmons C;	
XX	WPI; 2000-387421/33.	
DR	N-PSDB; AAA27407.	
XX	N-PSDB; AAA27407.	
PT	New polynucleotide encoding plant aminoacyl-tRNA synthase and the	

Aspergillus fumiga
Aspergillus fumiga
Propionibacterium
N. meningitidis va
N. gonorrhoeae am
Putative P. abyssi
Arabidopsis thalia
Candida albicans e
Human ORF2056
Drosophila melano
Novel human enzy
N. gonorrhoeae am
Drosophila melano
H. pylori GPO 169
Helicobacter pylor
Haemophilus influe
Streptococcus poly
Enterococcus faeca
E. coli cellular p
Enterococcus faeca
Lactococcus lactis
Listeria monocyog
Streptococcus poly
Streptococcus poly
Soybean glutamyl-t
S. pneumoniae type
C. pneumoniae prot
Human ORF protein
Staphylococcus aur
Glutamyl-tRNA synt
Staphylococcus aur
Streptococcus poly
Streptococcus poly
Streptococcus pneu
Salmonella typhi c
Streptococcus pneu
Glutamyl tRNA synt

PT encoded polypeptide, useful e.g. for regulating gene expression -
XX
XX Claim 11; Page 48-50; 74pp; English.
XX
XX The present sequence is the glutamyl-tRNA synthetase for corn. The enzyme
CC is an aminoacyl-tRNA synthetase (AARS). AARS charge (acylate) specific
CC tRNAs with amino acids for use in protein synthesis. Glutamyl-tRNA
CC synthetase therefore charges a glutamyl-tRNA with glutamate. Since this
CC enzyme has a crucial role in protein synthesis and therefore life, any
CC agent that inhibits or disrupts protein synthesis is likely to be toxic.
CC The present sequence could therefore be used as a basis for testing
CC whether the encoded aminoacyl-tRNA synthetase is sensitive to known
CC inhibitors or other chemicals and hence could be used in the discovery of
CC potential herbicides.
XX
XX

Sequence 715 AA;

Query Match 100.0%; Score 3734; DB 21; Length 715;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEALSFSDSPSIITICAKLVGLPLTINSLAAGSAPTQFASGESLHGVNPIITLXIA 60
DB 1 MEALSFSDSPSIITICAKLVGLPLTINSLAAGSAPTQFASGESLHGVNPIITLXIA 60
QY 61 RGASIASISGKNDIEFGHVEWLEVAAPFLISGSEFENACLFVDGFLASRTFLVGHGLTIA 120
DB 61 RGASIASISGKNDIEFGHVEWLEVAAPFLISGSEFENACLFVDGFLASRTFLVGHGLTIA 120
QY 121 DIAVWSNLAGIGORWESLRKSKYONLVWFNSIDSEYKALNEVVAFVGKRGIGKSPA 180
DB 121 DIAVWSNLAGIGORWESLRKSKYONLVWFNSIDSEYKALNEVVAFVGKRGIGKSPA 180
QY 181 PSLEKAYDSKDPAPAEVDLPQAKYKVCVRFAPSPSYLHIGAKALINKFYAERYOG 240
DB 181 PSLEKAYDSKDPAPAEVDLPQAKYKVCVRFAPSPSYLHIGAKALINKFYAERYOG 240
QY 241 RLIVRPDDTNPSSKESNEFENLKDIEFLGIKYDAVYTSOYFPKLMEMASLIIQOGKAY 300
DB 241 RLIVRPDDTNPSSKESNEFENLKDIEFLGIKYDAVYTSOYFPKLMEMASLIIQOGKAY 300
QY 301 IDDPKKEQMRKERMDGISRCRNNTVEENLSLMEKNVNGTRGMOCCVRGKLDMDPKKS 360
DB 301 IDDPKKEQMRKERMDGISRCRNNTVEENLSLMEKNVNGTRGMOCCVRGKLDMDPKKS 360
QY 361 LRDPVYRCNTDPHHRVGSKYKYVPTYPFACPFVDALSGVTHALRSSSEYHBRNAQYRIL 420
DB 361 LRDPVYRCNTDPHHRVGSKYKYVPTYPFACPFVDALSGVTHALRSSSEYHBRNAQYRIL 420
QY 421 QDMGLRVEIYEFSLNNVYTLSSKRLLMFVNKKVEDWTDPRPTVOGIVRGKLYEA 480
DB 421 QDMGLRVEIYEFSLNNVYTLSSKRLLMFVNKKVEDWTDPRPTVOGIVRGKLYEA 480
QY 481 LIQFLIOOGASRKNLMEWDLMTINKKIIPVCRHRAVLKDQVITLTNNGPSEBPVR 540
DB 481 LIQFLIOOGASRKNLMEWDLMTINKKIIPVCRHRAVLKDQVITLTNNGPSEBPVR 540
QY 541 ILPRHKKPEGAGKATTFANRIWLDYADAAAIKNGEETLMDGNALYKEIKVEGVTTE 600
DB 541 ILPRHKKPEGAGKATTFANRIWLDYADAAAIKNGEETLMDGNALYKEIKVEGVTTE 600
QY 601 LVGSLHLEGSVYTKTKITWLADIEBLVPLSVEFDYLSKKKLEDEDFLDNLNPTCRR 660
DB 601 LVGSLHLEGSVYTKTKITWLADIEBLVPLSVEFDYLSKKKLEDEDFLDNLNPTCRR 660
QY 661 EIPALGDANMRNIKGEIIOLEKRYRCDAFITSSSRPVULFAIPDRQOASLS 715
DB 661 EIPALGDANMRNIKGEIIOLEKRYRCDAFITSSSRPVULFAIPDRQOASLS 715

RESULT 2
AAG41411
ID AAG41411 standard; Protein; 716 AA.

XX
AC AAG41411;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 51519.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX
XX 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
XX
XX 23-MAR-1999; 99US-0125788.
XX
XX 25-MAR-1999; 99US-0126264.
XX
XX 29-MAR-1999; 99US-0126785.
XX
XX 01-APR-1999; 99US-0127462.
XX
XX 06-APR-1999; 99US-0128234.
XX
XX 08-APR-1999; 99US-0128714.
XX
XX 16-APR-1999; 99US-0129845.
XX
XX 19-APR-1999; 99US-0130077.
XX
XX 21-APR-1999; 99US-0130449.
XX
XX 23-APR-1999; 99US-0130510.
XX
XX 23-APR-1999; 99US-0130891.
XX
XX 28-APR-1999; 99US-0131449.
XX
XX 30-APR-1999; 99US-0132048.
XX
XX 04-MAY-1999; 99US-0132407.
XX
XX 05-MAY-1999; 99US-0132484.
XX
XX 06-MAY-1999; 99US-0132485.
XX
XX 06-MAY-1999; 99US-0132486.
XX
XX 07-MAY-1999; 99US-0132487.
XX
XX 11-MAY-1999; 99US-0132863.
XX
XX 14-MAY-1999; 99US-0134256.
XX
XX 14-MAY-1999; 99US-0134218.
XX
XX 14-MAY-1999; 99US-0134219.
XX
XX 14-MAY-1999; 99US-0134221.
XX
XX 14-MAY-1999; 99US-0134370.
XX
XX 18-MAY-1999; 99US-0134768.
XX
XX 19-MAY-1999; 99US-0134941.
XX
XX 20-MAY-1999; 99US-0135129.
XX
XX 21-MAY-1999; 99US-0135353.
XX
XX 24-MAY-1999; 99US-0135629.
XX
XX 25-MAY-1999; 99US-0136021.
XX
XX 27-MAY-1999; 99US-0136392.
XX
XX 28-MAY-1999; 99US-0136782.
XX
XX 01-JUN-1999; 99US-0137222.
XX
XX 03-JUN-1999; 99US-0137528.
XX
XX 04-JUN-1999; 99US-0137502.
XX
XX 07-JUN-1999; 99US-0137724.
XX
XX 08-JUN-1999; 99US-0138094.
XX
XX 10-JUN-1999; 99US-0138540.
XX
XX 10-JUN-1999; 99US-0138847.
XX
XX 14-JUN-1999; 99US-0139119.
XX
XX 16-JUN-1999; 99US-0139452.
XX
XX 16-JUN-1999; 99US-0139453.
XX
XX 17-JUN-1999; 99US-0139492.
XX
XX 18-JUN-1999; 99US-0139454.
XX
XX 18-JUN-1999; 99US-0139455.
XX
XX 18-JUN-1999; 99US-0139456.
XX
XX 18-JUN-1999; 99US-0139457.
XX
XX 18-JUN-1999; 99US-0139458.
XX
XX 18-JUN-1999; 99US-0139459.
XX
XX 18-JUN-1999; 99US-0139460.
XX
XX 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 26-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144337.
PR 19-JUL-1999; 99US-0144338.
PR 19-JUL-1999; 99US-0144339.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147199.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149425.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151068.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query March 71.2%; Score 2657; DB 21; Length 716;
Best Local Similarity 69.0%; Pred. No. 1.6e-226;
Matches 491; Conservative 106; Mismatches 107; Indels 8; Gaps 5;

QY 5 LSFESKSPPIISIIICAAKLVGLPTINHSILAASAPTIQFASGESLHGVPPIILYIARGA- 63
3 LSFPPESPPLSVIYALSLASPVITIDSSAAATVPSPFSDGRLNGATVLRVGRSAK 62
QY 64 SIASLSKNDIEFGHVVEWLEYPATPLSGSEFFENACLFVDGFLASRFFLVGHGLTIADIA 123
63 KLDPFYGNNAFDSQIDBWDYASVFSSGSEFFENACGVRDXYLESSTFLVGHSLSIADVA 122
QY 124 VMSNLAGIGORWESLRSKSKYKONLVRFNNSIDSEYKXALNVAAPYKRGIGKS-PAPS 182
123 IWSALAGIGORWESLRSKSKYKONLVRFNNSIDSEYKXALNVAAPYKRGIGKS-PAPS 181
QY 183 LKEKRVH----DSKDPSPAPEVDLPQAKYGVKCVRFAPRPSGYLHIGHAKAALNXYFAERY 238
182 SKDSQAVKGDGQKGRPEVDLPQAKYGVKCVRFAPRPSGYLHIGHAKAALNXYFAERY 241

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 71.2%; Score 2657; DB 21; Length 719;
Best Local Similarity 69.0%; Pred. No. 1,66-226;
Matches 491; Conservative 106; Mismatches 107; Indels 8; Gaps 5;

QY 5 LSFESKSPPIISICAAKLVGLPLTINSHLAGAPTLQFASGESLHGVNPIITYARCA- 63
DB 6 LSFEPSEPLSLVIALSLASAPVTIDSSAAATVPSEVFSGKRLKGATVLLRYVGRSAK 65
QY 64 SIASLSGKNDIEFPHVWEMLEAPPTLSGSFEENACLVFDGFLASRTFLVGHGLTIADIA 123
DB 66 KLDPFYGNNAFDSQIDIEWVDYASVFSGSSEFFENACGVDRYALSSSTFLVGHSLSIADIVA 125
QY 124 VMSNLAGIGQRWESLRSKSKYQNLVRFNFSIDSEYKEALNEVNAFYGKRGIGKS-PAPS 182
DB 126 IWSALAGTQGMESLRSKSKYQSLVRFNFSILDEYSEVNLKVATLVY-KKSGKRVAPAK 184
QY 183 LKEKVH---DSKPSAPEVDLPEAKVGVKCVRPAPESGYLHIGHAAVALLNKYFARY 238
DB 185 SKDSQAVKGDQKGRPEVDLPEALIGKVLRAPAPESGYLHIGHAAVALLNKYFARY 244
QY 239 QGRLLVFPDPTNPKESNEFENLKDIENTIGYDANTYSDVPEPKIMEAESLKGCK 298
DB 245 QGEVIVRPDDTNPAKESNEFVDNLVKDITGITYEKYTSDFPELMDNAEKLMRGK 304
QY 299 AYIDTPEKQMRKRMGIESRCRNNTVEENLSMKEMVNGTERGMQCCVYKGLDMODPN 358
DB 305 AYVDTPREQOKARMGIDSKCNHSEVENLAKMKEMIASERGLQCCVYKRMODPN 364
QY 359 KSLRDPVYYRCNTDPHHRVSGKYVPTYDFACPFDVALJEGVTHALRSSEVHDAQAQYR 418
DB 365 KAMDVPYYRCNPMSHRIGDKYKITYDFACPFDVLSGITHALRSSEVHDAQAQYF 424
QY 419 ILDMGLRVEITYEFSRLNMYTTLISKRKLMPQNKKEVEMTPRPFTVGGIYRGLKV 478
DB 425 VLEDMGLRQVQLYEFSRLNMYTTLISKRKLMPQNKKEVEMTPRPFTVGGIYRGLKI 484
QY 479 EALIOFLIOGASAKNLMEMDKLMTINKKIIDVCAKHTAVLDDORVIFLLTNGPEBP 538
DB 485 EALIOFLIOGASAKNLMEMDKLMTINKKIIDVCPRHIAVVERVYVLLFLIOGPEBP 544
QY 539 VRIILPRHKKFEGACKKATTFANRIMLDYADAAANKGEVTLMDGNAIVKEI-KVESGV 597
DB 545 VRMI PKHKKFEGACKKATTFKSIMLEBADAASISVGEVTLMDGNAIVKEITKDEGR 604
QY 598 ITELVGSLHEGSKYTKTKLTWLADEBELVPLSLVEFDYILSKKKLEBDEDFLDNLNFC 657
DB 605 VTALSGVNLQGSYKTKTKLTWLPDNLVNLTLTEBDYILITKKKLEBDEDEVDVFNPN 664
QY 658 TRREIPALGDANMNRINRGEIIOLEKGYVACDAPTRSSKPVVLLIAPDGR 709
DB 665 TKKETTLALGDSNMNRNLKCGDYIOLEKGYFCADVFPVSSKPVVLLIAPDGR 716

RESULT 4
AAG41409

```
ID  AAC41409 standard; Protein; 748 AA.
XX
XX  AAC41409;
AC
XX  18-OCT-2000 (first entry)
DT
XX  18-OCT-2000 (first entry)
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 51517.
XX
XX  Protein identification; signal transduction pathway; metabolic pathway;
KM  hybridisation assay; genetic mapping; gene expression control; promoter;
XX  termination sequence.
OS  Arabidopsis thaliana.
XX  EP1033405-A2.
XX  EP1033405-A2.
XX  06-SEP-2000.
PD
XX  25-FEB-2000; 2000EP-0301439.
PF
XX  25-FEB-1999; 99US-0121825.
XX  05-MAR-1999; 99US-0123180.
XX  09-MAR-1999; 99US-0123548.
XX  23-MAR-1999; 99US-0125788.
XX  25-MAR-1999; 99US-0126264.
XX  29-MAR-1999; 99US-0126785.
XX  01-APR-1999; 99US-0127462.
XX  06-APR-1999; 99US-0128234.
XX  08-APR-1999; 99US-0128714.
XX  16-APR-1999; 99US-0129845.
XX  19-APR-1999; 99US-0130077.
XX  21-APR-1999; 99US-0130044.
XX  23-APR-1999; 99US-0130510.
XX  28-APR-1999; 99US-0130891.
XX  30-APR-1999; 99US-0131449.
XX  30-APR-1999; 99US-0132048.
XX  04-MAY-1999; 99US-0132407.
XX  04-MAY-1999; 99US-0132484.
XX  05-MAY-1999; 99US-0132485.
XX  06-MAY-1999; 99US-0132486.
XX  07-MAY-1999; 99US-0132487.
XX  11-MAY-1999; 99US-0132863.
XX  14-MAY-1999; 99US-0134256.
XX  14-MAY-1999; 99US-0134218.
XX  14-MAY-1999; 99US-0134219.
XX  14-MAY-1999; 99US-0134221.
XX  14-MAY-1999; 99US-0134370.
XX  18-MAY-1999; 99US-0134766.
XX  19-MAY-1999; 99US-0134941.
XX  20-MAY-1999; 99US-0135124.
XX  21-MAY-1999; 99US-0135353.
XX  24-MAY-1999; 99US-0135629.
XX  25-MAY-1999; 99US-0136021.
XX  27-MAY-1999; 99US-0136392.
XX  28-MAY-1999; 99US-0136782.
XX  01-JUN-1999; 99US-0137222.
XX  03-JUN-1999; 99US-0137528.
XX  04-JUN-1999; 99US-0137502.
XX  07-JUN-1999; 99US-0137724.
XX  08-JUN-1999; 99US-0138594.
XX  10-JUN-1999; 99US-0138549.
XX  10-JUN-1999; 99US-0138847.
XX  14-JUN-1999; 99US-0139119.
XX  16-JUN-1999; 99US-0139452.
XX  17-JUN-1999; 99US-0139453.
XX  17-JUN-1999; 99US-0139492.
XX  18-JUN-1999; 99US-0139454.
XX  18-JUN-1999; 99US-0139455.
XX  18-JUN-1999; 99US-0139456.
XX  18-JUN-1999; 99US-0139457.
XX  18-JUN-1999; 99US-0139458.
XX  18-JUN-1999; 99US-0139459.
XX  18-JUN-1999; 99US-0139460.
XX  18-JUN-1999; 99US-0139461.
XX  18-JUN-1999; 99US-0139462.
XX  18-JUN-1999; 99US-0139463.
XX  18-JUN-1999; 99US-0139750.
XX  18-JUN-1999; 99US-0139763.
XX  21-JUN-1999; 99US-0139817.
XX  22-JUN-1999; 99US-0139899.
XX  23-JUN-1999; 99US-0140353.
XX  23-JUN-1999; 99US-0140354.
XX  24-JUN-1999; 99US-0140695.
XX  28-JUN-1999; 99US-0140823.
XX  29-JUN-1999; 99US-0140991.
XX  30-JUN-1999; 99US-0141287.
XX  01-JUL-1999; 99US-0141842.
XX  01-JUL-1999; 99US-0142154.
XX  02-JUL-1999; 99US-0142055.
XX  06-JUL-1999; 99US-0142390.
XX  08-JUL-1999; 99US-0142803.
XX  09-JUL-1999; 99US-0142920.
XX  12-JUL-1999; 99US-0142977.
XX  13-JUL-1999; 99US-0143542.
XX  14-JUL-1999; 99US-0143624.
XX  15-JUL-1999; 99US-0144005.
XX  16-JUL-1999; 99US-0144085.
XX  16-JUL-1999; 99US-0144086.
XX  19-JUL-1999; 99US-0144325.
XX  19-JUL-1999; 99US-0144331.
XX  19-JUL-1999; 99US-0144332.
XX  19-JUL-1999; 99US-0144333.
XX  19-JUL-1999; 99US-0144334.
XX  19-JUL-1999; 99US-0144335.
XX  20-JUL-1999; 99US-0144352.
XX  20-JUL-1999; 99US-0144632.
XX  20-JUL-1999; 99US-0144884.
XX  21-JUL-1999; 99US-0144814.
XX  21-JUL-1999; 99US-0145086.
XX  21-JUL-1999; 99US-0145088.
XX  22-JUL-1999; 99US-0145085.
XX  22-JUL-1999; 99US-0145087.
XX  22-JUL-1999; 99US-0145089.
XX  22-JUL-1999; 99US-0145192.
XX  23-JUL-1999; 99US-0145145.
XX  23-JUL-1999; 99US-0145218.
XX  23-JUL-1999; 99US-0145224.
XX  26-JUL-1999; 99US-0145276.
XX  27-JUL-1999; 99US-0145913.
XX  27-JUL-1999; 99US-0145918.
XX  27-JUL-1999; 99US-0145919.
XX  28-JUL-1999; 99US-0145951.
XX  02-AUG-1999; 99US-0146386.
XX  02-AUG-1999; 99US-0146388.
XX  02-AUG-1999; 99US-0146389.
XX  03-AUG-1999; 99US-0147038.
XX  04-AUG-1999; 99US-0147204.
XX  04-AUG-1999; 99US-0147302.
XX  05-AUG-1999; 99US-0147132.
XX  05-AUG-1999; 99US-0147260.
XX  06-AUG-1999; 99US-0147303.
XX  06-AUG-1999; 99US-0147416.
XX  09-AUG-1999; 99US-0147493.
XX  09-AUG-1999; 99US-0147935.
XX  10-AUG-1999; 99US-0148171.
XX  11-AUG-1999; 99US-0148319.
XX  12-AUG-1999; 99US-0148341.
XX  13-AUG-1999; 99US-0148565.
XX  13-AUG-1999; 99US-0148684.
XX  16-AUG-1999; 99US-0149358.
XX  17-AUG-1999; 99US-0149175.
XX  18-AUG-1999; 99US-0149426.
XX  20-AUG-1999; 99US-0149722.
XX  20-AUG-1999; 99US-0149723.
XX  20-AUG-1999; 99US-0149929.
XX  23-AUG-1999; 99US-0149902.
```

```
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 71.2%; Score 2657; DB 21; Length 748;

Best Local Similarity 69.0%; Pred. No. 1.7e-226; Mismatches 107; Indels 8; Gaps 5;

Matches 491; Conservative 106; Mismatches 107; Indels 8; Gaps 5;

```
OY 5 LFSKDSPPISIIICAKLVGLPLTINHSIAGSAPTLQFASGSESHGVNPIIILYARGA- 63
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 35 LSPFSPSPPLVIVALSASAPVITDSSAAATTVSPFSPDRKLNKATVLIRYGRSAK 94
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 64 SIASISGKNIDIEFGHVEWLEVAPTFLSGSEFENACLFVDGFLASRTFLVGHGLTIADIA 123
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 95 KLDPFYGNNAFDSQIDIEWVDYASVSSGSEFENACGRVDKYLESSSTFLVGHSLIADVA 154
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 124 VMSNLAGIGORWESIRKSKKONTVRWFNSIDSEKELNEVVAFVGKRGIGKS-PAPS 182
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 155 IWSALAGIGORWESIRKSKKQSLVRWFNSIIDSEYVINKLAVYV-KKSGKRVAAPK 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 183 LKEKVH---DSKSDSAPEVDLPGAKVGKVCVRPAPEPSGYLHIGAKAALLNKYFARY 238
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 214 SKDSQAAVKGDDQDKGKEVDLPEAEIGKVKLRAPAPESGYLHIGAKAALLNKYFARY 273
OY QGRILVFPDPTNPKESNEFENLTKDIETLGIYADATYTSDFPKLMEASLTKGK 298
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 274 QGEVIVFPDPTNPKESNEFVDNLVKDITIGITKEVYTSDFPELMDPAEKLMRGK 333
OY AYIDDTPEKQMRKEMDIESRCRNVTVEENISLMEKWNTERGMOCCVAKLDMODPN 358
Db 334 AYVDDTPEQWQKERMGIDSKCRNHSVEENLKLMEKMIAGSEGLQCCVAKGFMODPN 393
OY KSLDPVYTRCNTDPHHRVSGKIVYPTYPACFPVDALBSVTALRSSEYHDRAQYTR 418
Db 394 KMRDPVYTRCNPMSHRIIDGKXIPTYDPAFCFVDSLBSITHALSSSEYHDRAQYFK 453
OY ILDMGRLRVEIYFESRLNMYTLLSKRKLMPFONKKEVMTDPRPFTVGIYRGLKY 478
Db 454 VLEDMLRQVOLYFESRLNMYTLLSKRKLMPFQGLVDMDDPRPFTVGIYRGLKI 513
OY EALIQFLIQGASQNLNLMEMDKLMTINKKIIDPVCARHTAVLRDORYFTLTNGPEBP 538
Db 514 EALIQFLIQGASQNLNLMEMDKLMSINKRIIDPVCPRHTAVVARRVLTFTDGPDEPF 573
OY VRIIPRHKKEFGAKKATTPANRIWLDYDAALNKGEVTLMDGNAIVKEI-KVESGV 597
Db 574 VRMI PKHKKEFGAGEKATTFPKSIWLEBDAASISVGEVTLMDGNAIVKEITDDEGR 633
OY ITTELVEGLHLEGSVKTTRKLTITWLADI BELVPLSLVEEDVYLSKKLEEDDFLDNIPC 657
Db 634 VTALSGVNLQGSVKTTRKLTITWLADITNELVNLTLTEFDYITTKKLEDDVDVFNPN 693
OY TRREIPALGDAMNENIKRGEIIQLEKGYVRCDAPIRSGKPVVLAIPDGR 709
Db 694 TKKETLALGDSNMNMLKCGDIQLEKGYRFDVDFPKVSSRPYLFSLPDGR 745
```

RESULT 5

ABP73689
ID ABP73689 standard; Protein, 725 AA.

AC ABP73689;

XX 30-JAN-2003 (first entry)

XX Candida albicans essential protein SEQ ID NO 7526.

XX Fungus, yeast; tetracyclin, promoter; GRACB strain; bioynthesis;

XX signal transduction; DNA replication; cell division; growth;

XX proliferation; Candida albicans; fungicide; antifungal.

XX Candida albicans.

XX WO200253728-A2.

XX 11-JUL-2002.

XX 26-DEC-2001; 2001WO-US49486.

XX 29-DEC-2000; 2000US-259128P.

XX 20-FEB-2001; 2001US-0792024.

XX 22-AUG-2001; 2001US-314050P.

XX (ELIT-) ELITRA PHARM INC.

XX Roemer T, Jjiang B, Boone C, Bussey H, Ohlsen KL;

XX WPI; 2002-566694/60.

XX N-PSDB; ABZ32239.

XX Constructing strains for identifying gene products as effective targets
XX for therapeutic intervention, by inactivating in the strain one allele
XX of a gene and placing other allele of the gene under conditional
XX expression -


```

Qy 177 KSPAPSLKE---KVHDSKDPAPAEVLDPAKVGKVCVAPAPSPGYLHIGAKAALLNKY 233
Db 171 RSPQSSKEQTPAKTERKO-ECKFVDLPQAEWGKVVAPFPASGYLHIGAKAALLNQY 229
Qy 234 FAERYQGLIYRFPDTPNSKESNEFEVNLKDIEFLGIKYDAVYTSYFPKLMEMASL 293
Db 230 YALAFQGLIYRFPDTPNAKEVEFENILDLDELQIKPDPFTTSYFDLMIDYCVRL 289
Qy 294 IKOGKAYIDTPKEOMRKERMDGIESRCRNNTVEENLSLMEKMNVTGERGMOCCVRGLD 353
Db 290 IKESKAYIDTPPEOMKLEREQRVESANRNSVERNLSLMEBVAWGSEKGYICVRAKID 349
Qy 354 MODPKSLRDPVYRCNTDPHHRVGSKKYVPTYDFACFPYDALEGVTHALRSSSEYHNRN 413
Db 350 MSPFNGCMRDPTRYCKNEPHPRTGTXKVPYTOFACPIYDALENVHTLRTTEYHRD 409
Qy 414 AQYRILODMGLRVEIYEFSLNMYTLTKSKRLMWONKVEDWTDPREPTYQGIYR 473
Db 410 DQFWFIDALTKRKRYISYSRLNMTNTVLTKSKRLTWFSVDSGLVDGMDPPFPYRGIIR 469
Qy 474 RGLKVEALIOFLIOGASKNLNMEMWDLMTINKKIIDPVCAHRTAVLKDORVIFTLNG 533
Db 470 RGMVTEGLKEPILIAQSSKSVFPMWMDKIMFNKKVIDIPAPRYTALSKERKVIYNVAGA 529
Qy 534 PEEFPVRI-LPRHKKFEGAGKATTFANRIMLDYADAANKGEBEVTLMGNAIVKEI- 591
Db 530 KVE---RIQVSVHPRDESLGKKTVLGPRYIYIDYDADALKEGENATINMGNIIRKVN 586
Qy 592 KVESGVITELVGLHLEGSVTKTKITWLA--DIEELVPLSLVEFDYLSKKLEEDE 648
Db 587 KDSAGNITSVDAALNLEKDKFPKTLKLTWLAVEDPSAPYPTFCYVFNIISKAVLGKDE 646
Qy 649 DELNINPCTREIIPAGDAMRNKRGEEILOERKGYRCDAPPIRS-----KPVVL 702
Db 647 DFKOPIGHIKTDEVPMLGDPBLKCKCKGDIILOQRGFKVDAVAPPSGYTNVSPPTVL 706
Qy 703 FAIPDG 708
Db 707 FSIIDG 712

RESULT 7
ID AAU04349 standard; Protein: 1512 AA.
XX AC AAU04349;
XX DT 23-OCT-2001 (first entry)
XX DE Mammalian toxicological response marker protein #1.
XX KM Mammalian toxicological response marker; antigen; antibody; agonist.
XX OS Homo sapiens.
XX PN WO200136684-A2.
XX PD 25-MAY-2001.
XX PE 16-NOV-2000; 2000WO-US31743.
XX PR 19-NOV-1999; 99US-0443184.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Cunningham MJ, Zweiger GB, Kaser MR, Panzer SR, Seilhamer JJ;
XX PI Yue H, Baughn MR, Azimzai Y, Lal P;
XX DR WPI: 2001-355646/37.
XX PT Novel mammalian nucleic acid molecules whose levels are up regulated or
down regulated following treatment with a toxic compound, useful for
detecting metabolic and toxicological responses and in monitoring drug

```

```

PT action -
XX Claim 9; Page 56-59; 89pp; English.
PS The sequence represents a novel mammalian protein encoded by a
XX nucleic acid molecule whose levels are up regulated or down regulated
CC following treatment with a toxic compound. Polynucleotide sequences
CC complementary to the sequences of the invention are useful for
CC preventing a toxicological response by acting against one or
CC more up-regulated nucleic acid molecules. An agonist identified by the
CC above method is also useful for preventing a toxicological response by
CC initiating transcription of a gene comprising a down regulated nucleic
CC acid molecule of the invention. Proteins encoded by the nucleic acids of
CC the invention are useful for producing antibodies. The nucleic acid
CC molecules are useful for producing an animal model system. The nucleic
CC acids immobilised on a substrate as hybridisable array element in a
CC microarray format may be used to characterise gene expression patterns
CC associated with novel compounds to elucidate any toxicological responses,
CC or to monitor the effects of treatments during clinical trials or therapy
CC where metabolic response to toxic compounds may be expected.
CC The nucleic acids are useful for various hybridisation technologies and
CC is useful for designing hybridisation probes. The nucleic acid molecule
CC or its fragment, or a protein encoded by the nucleic acid molecule may be
CC used to purify a ligand from a sample.
XX
SQ Sequence 1512 AA;
Query Match 40.3%; Score 1506; DB 22; Length 1512;
Best Local Similarity 42.2%; Pred. No. 8.2e-124;
Matches 305; Conservative 146; Mismatches 232; Indels 40; Gaps 9;
Qy 4 ALSPKSDSPSITICAKVLGVLPLTNHSLAAGSAPLQFASGSLMGVNIILYIRGA 63
Db 5 SLTVNSGDPPLGALAAVEHVAVDVSI--SVEEGENILHSENVIFTDVNSILYLRKVA 62
Qy 64 SIASISGKNDIEFHVVEWLEIYAPFLSGE-FENACLFVDGFLASRTFLVGHGLTIADI 122
Db 63 TTAGLVGSNLMHEHIDHWLEFSAITKSSCDSFSTINELNHCISLRTYLVGNSLSLADL 122
Qy 123 AVWSNLAGIQRMESLRSKKYONLVRFNSIDSEYKALNEVVAAPVKKRGICKSPAPS 182
Db 123 CVMATLKGNAMWQOLQKKAPVHVKKRMFGLAQ--QAFQSV-----G 164
Qy 183 LKEKVHDSKDPSPAE-----VDLPKAVKVCVAPAPSPGYLHIGAKAALLNKYF 234
Db 165 TKMDVSTTKARVAPEKKQDVGVFELPQAEKGVTFPPASGYLHIGAKAALLNQHY 224
Qy 235 AERYQGLIYRFPDTPNSKESNEFEVNLKDIEFLGIKYDAVYTSYFPKLMEMASL 294
Db 225 QVNFKGLIYRFPDTPNEKEKEDEKYLIEDVAMLHKPQDFTTSDHFEITMYAKRLI 284
Qy 295 KQKRAYIDTPKEOMRKERMDGIESRCRNNTVEENLSLMEKMNVTGERGMOCCVRGLDM 354
Db 285 QEGKAYIDTPPEOMKLEREQRVESANRNSVERNLSLMEBVAWGSEKGYICVRAKIDM 344
Qy 355 QDPKSLRDPVYRCNTDPHHRVGSKKYVPTYDFACFPYDALEGVTHALRSSSEYHNRN 414
Db 345 SSNNGCMRDPTRYCKIQPHPRGKNKYVPTYDFACPIYDASLEGVTHALRTTEYHRDE 404
Qy 415 QYVRILODMGLRVEIYEFSLNMYTLTKSKRLMWONKVEDWTDPREPTYQGIYR 474
Db 405 QFYVITIALGIRKRYIYEFSLNMYTLTKSKRLTWFSVDSGLVDGMDPPFPYRGIIR 464
Qy 475 RGLKVEALIOFLIOGASKNLNMEMWDLMTINKKIIDPVCAHRTAVLKDORVIFTLNGP 534
Db 465 GMTVEGLKEPILIAQSSKSVFPMWMDKIMFNKKVIDIPAPRYTALSKERKVIYNVBEAQ 524
Qy 535 EEPVRI-LPRHKKFEGAGKATTFANRIMLDYADAANKGEBEVTLMGNAIVKEI-KV 593
Db 525 EE-MKEVAKPKNPEVGLRPVWSPVFIEGDAETFSGSENVTFIMGNLNTKLIHKN 582
Qy 594 ESGVITELVGLHLEGSVTKTKITWLAIDIEELVPLSL--VEFDYLSKKLEEDDFL 651

```

Db 583 ADGKIISLDANKLENKDYKTKTKVTWLAETHALPIVICVTEYHILITKPVLGKDEDFK 642
QY 652 DNLNPTCTREIPALGDANMRNIRKGEIITQLERKGYRCD-----APFIRSSKPVLPFAI 705
Db 643 QVNNKNSGHEBLMLGDPLCKDKLKKGDIQLORRGFPICDPYEPVSPYSCKEAPCVLIYI 702
QY 706 PDG 708
Db 703 PDG 705

RESULT 8
AAM78732 standard; Protein; 1512 AA.
AAM78732;
AA78732;
06-NOV-2001 (first entry)
Human protein SEQ ID NO 1394.
Human; cytokine; cell proliferation; cell differentiation; gene therapy;
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
tissue growth factor; immunomodulatory; cancer; leukaemia;
nervous system disorder; arthritis; inflammation.
Homo sapiens.
MO200157190-A2.
09-AUG-2001.
05-FEB-2001; 2001WO-US04098.
03-FEB-2000; 2000US-0496914.
27-APR-2000; 2000US-0560875.
20-JUN-2000; 2000US-0598075.
19-JUL-2000; 2000US-0620325.
01-SEP-2000; 2000US-0654936.
15-SEP-2000; 2000US-0663561.
30-OCT-2000; 2000US-0693325.
30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
PI Zhao QB, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
XX N-PSDB; AAK51865.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 3654-3657; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
SQ Sequence 1512 AA;

Query Match 40.2%; Score 1502; DB 22; Length 1512;
Best Local Similarity 42.0%; Pred. No. 1.96-123;
Matches 304; Conservative 147; Mismatches 232; Indels 40; Gaps 9;
QY 4 ALSFSDSPISITICAAKLVGLPTLNHSLAAGAPLQFASGSLGVNPIIYIARGA 63
Db 5 SLTVNSGDPPLGALLAVEHVADVSI--SVLEGENTLHSENVIFPDVNSILNARVA 62
QY 64 SIASLGRNDIEFGHVVEWLEAPTFLSGSF--FENACLFDVDFLASRTFLVGHGLTADI 122
Db 63 TTAGLYGSNLMETHREIDHMLEFSATKLSGCSFSTINELNHCLSLRTYLVGNSLADL 122
QY 123 AVMSNLAGIGORMSLSKSKYQNLVWVENSIDSEYKALNEVVAAPFGKRGIGKSPRS 182
Db 123 CVMATLKGNAAMQOLKQKAPVHVKKWFFLEAQ--QAFQSV-----G 164
QY 183 LKEKYHDSKDSAPE-----VDLPQAKYGVCFVFPAPSPGYLHGHAKAALLNRYF 234
Db 165 TKMDVSTTKARVAPEKKQDVGFELFGLAEMGKTVAFPPBASGYLHGHAKAALLNRY 224
QY 235 AERYQGLIVAFDDTNPSSKESNEFVENLADIEITLGIKYDAVYTSYDFPYLMEMASLI 294
Db 225 QVNFKGLIMFPDPTNPEKKEDFEKYLIEDVAMLHKPDQFTYSDHFEIMKYAEKLI 284
QY 295 KQGAIVDDTPKEQMKERKNDGIESRCRANTVEENLSLMKMNANGTERGMQCCVRGLDM 354
Db 285 QEGRAVYDDTPAEQMKERQRISSKRRKPIEKNLQWMEEMKKSQFGQSCIRAKIDM 344
QY 355 QDPNKSJLDPVYVRCNTDPHHRVGSKYKVPYFPACFPVDALGVTALRSSEYHDNA 414
Db 345 SSNNGCRNDPFLYCKIOPHRTGKKNVPTTFPACPIYDSIGVTHALRTTEYHDE 404
QY 415 QYRILQDMGLRVEIYEFSLNMYTLLSKRKLMPVQNKVEDWTDPRFPYQGIYVR 474
Db 405 QFYWIEALGIRKQYIWEYSRLNNTVLSRKLTWFMNEGLVGMDDPREPYRGVLR 464
QY 475 GLKYPALIOFLQOGASGNLMLMWDKMTINKKIIDPVCARHNAVLDQGVVIFLTNGP 534
Db 465 GMTVEGLKQFIQAOGSSRSVNMEMDKIMAFNKKVIDPVARVYALLKKEVIPPVNVBAQ 524
QY 535 EEPFVRLPRHKKEGAGKATTEFANRIMLDYADAAMINKSEVTLMDMGNAIYKEI-KV 593
Db 525 EE--MKVAKHPKPEVGLKPVWYSPKVIEGADAEFTSEGEWTFIMGNLNTIKHN 582
QY 594 ESVGITBELVGEHLLEGSVKTKTKITMTLADIEELVPLSL--VEEDYLISKKLEDEDFL 651
Db 583 ADGKIISLDANKLENKDYKTKTKVTWLAETHALPIVICVTEYHILITKPVLGKDEDFK 642
QY 652 DNLNPTCTREIPALGDANMRNIRKGEIITQLERKGYRCD-----APFIRSSKPVLPFAI 705
Db 643 QVNNKNSGHEBLMLGDPLCKDKLKKGDIQLORRGFPICDPYEPVSPYSCKEAPCVLIYI 702
QY 706 PDG 708
Db 703 PDG 705

RESULT 9
AAM79716 standard; Protein; 1550 AA.
AAM79716;
06-NOV-2001 (first entry)
Human protein SEQ ID NO 3362.
Human; cytokine; cell proliferation; cell differentiation; gene therapy;
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
tissue growth factor; immunomodulatory; cancer; leukaemia;
nervous system disorder; arthritis; inflammation.
XX

OS Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
PI Zhao Q, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX N-PSDB; AAK52849.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 20; Page 312; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX Sequence 1550 AA;
SQ
Query Match 40.0%; Score 1493; DB 22; Length 1550;
Best Local Similarity 41.9%; Pred. No. 1,2e-122;
Matches 303; Conservative 147; Mismatches 233; Indels 40; Gaps 9;
QY 4 A L S F S K D S P P I S I C A K L V G L P L T I N H S L A G S A P T L Q F A G S E L H G V N P I I L Y I A R G A 63
DB 43 S L T V N S G D P P I G A L L A V E H V K D V S I - S V E R G K E N I I H V S E N V I F T V N S I R L A R A V A 100
QY 64 S I A S I S G K N D I E F G V N V E M L E A P P T L S G S E F E N A C I F V D G F L A S R F L Y G H G I T I A D I 122
DB 101 T T A G I Y G S N L M E H R E I D I M L E F S A T K L S C D S F P T S T I N E L N H C I S L R T Y L V N S I L S L A D L 160
QY 123 A V M S N L A G I G O R M E S L R K S K K Y O N L V R W F N S I D S E Y K E A L M E V N A A F A P G K G I G S P A P S 182
DB 161 C V M A T L K N A M A Q E D L K O K A P A V H K M F G L E A Q - Q A F Q S V - - - - - G 202
QY 183 L K E K V H D S K D S A P P - - - - - V D L P G A K V G K V C R P A P E S P G Y L A H I G H A K A A L L I N K Y F 234
DB 203 T K M D V S T T K A R A P E K K O D V G F V E L P G A E M G K V T V R P P E A S G Y L H I G H A K A A L I N H Y 262
QY 235 A E R V O G R I V A P F D D T N P S K E S E F E V E N L K D I E T I G I T A D V A T T Y S D Y P P K L M E A S E L I 294
DB 263 Q V N F G K L I M P F D D T N P E K E K E D F E K V I L E D V A M A H I P R D P T Y T S D H F E T I M K Y A E K L I 322
QY 295 K O G K A Y I D P T P R E Q R K E R M D G I E S R C A N N T V E E N L S L M K E V N G T E R G M Q C V R G K L D M 354

DB 323 Q E G R A Y V D D T P G E O I K A R E O R I E S K H R K N P I E K N L Q M W E E K K G S Q F G H S C C L F A K I D M 382
QY 355 Q D P M K S L B D P V Y R C N D P H H R V G S K Y V P T Y P A C P F V A L B G V T A L S S E V H D N A 414
DB 383 S S N G C M D P T L Y C K I Q P H P R T G K Y N V P T Y P A C I V D S I E G V T A L T T E Y H D R D E 442
QY 415 Q Y R I L O D M G L R V E I Y E F S R L N M V Y T L S K R K L M F V O N K V E D M T P R P P T V O G I V R 474
DB 443 Q P Y I T E A L G I R K P Y I W S Y S L N I N N T V L S R K L T M F N E B G L V D C M D P R P P T V R G V L R 502
QY 475 G L K V E A L I Q P L L O G A S K N L M E M D K M T I N K K I I D P V C A R H T A V L K D Q R V I F T L T N G P 534
DB 503 G M T V E G L K O F I A O G S S R S V N M E M D K I W A R N K K V I D P V A R R Y A L L K E V I P A V N V P A Q 562
QY 535 E E P F R I L P R K K E F G A K K A T T P A N R I W L D Y A D A A A I N K E E V T L M W G N A I V E I - K V 593
DB 563 E E - N K E V A K H P K D P E V G L K P W Y S P K V I E G A D A E T P S E G E M V F I M W G I N I T K I H K N 620
QY 594 E S G V I T E L V G E L H E G S V K T T K L I T W L A D I E E L V P L S L - - V E P D Y L S K K K L E E D E D P L 651
DB 621 A D G K I S I D A K L N L E N D Y K K T T K Y T W L A E T T H A L P I V I C V T Y E H L T K P Y L G D E D P K 680
QY 652 D N L N P C T R R E I P A L G D A N M R N I K G E I I Q L E R K G Y R C D - - - - - A P P I R S K P V L P A I 705
DB 681 Q V N M K S K H E E L M L G D P C L K D L K K G D I Q L G R G P F I C D Q E Y E P V S P S C K E A P C V L I Y I 740
QY 706 P D G 708
DB 741 P D G 743
RESULT 10
ABU25717
ID ABU25717 standard; Protein; 715 AA.
XX ABU25717;
AC 16-APR-2003 (first entry)
XX 16-APR-2003 (first entry)
XX Aspergillus fumigatus essential gene protein #375.
XX DE Aspergillus fumigatus essential gene protein #375.
XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
XX KW cancer; contamination; biofilm; antibody; immune response.
XX OS Aspergillus fumigatus.
XX WO200286090-A2.
XX 31-OCT-2002.
XX 23-APR-2002; 2002WO-US13142.
XX 23-APR-2001; 2001US-285697P.
XX 27-APR-2001; 2001US-287066P.
XX 05-JUN-2001; 2001US-295890P.
XX 09-JUL-2001; 2001US-303899P.
XX 31-AUG-2001; 2001US-316362P.
XX (ELIT-) ELITRA PHARM INC.
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX New purified or isolated nucleic acids of essential genes of
XX Aspergillus fumigatus, useful for treating or preventing infections by
XX A. fumigatus, or for treating a non-infectious disease in a subject
XX e.g. cancer -
XX Disclosure; Page -; 175pp; English.
XX The invention relates to novel purified or isolated nucleic acids of

Query Match	39.4%	Score 1471.5	DB 24	Length 715
Best Local Similarity	43.0%	Pred. No. 2.9e-121		
Matches	310	Conservative 128	Mismatches 234	Indels 49
			Gaps	14
QY	4	ALSPKSDSPSIICAAKLVGLPTLTINSLAASAPTLQF--ASGESLHGVPILLYAR	61	
DB	26	SINEARPPSVISITVEDTRAV-----LRGDKRAVVGYSAGSNPIRGLINAVQELRK	76	
QY	62	GASISLSGKND-IEFGHVEVLETPF--LSGSEFENACLFVDFGLASRFLVGHGTT	118	
DB	77	--DPEPLNSKOEKLE---NEWLSQLEAFAPLDFALPDELQRLTHLIRSFVVGVALS	130	
QY	119	IADIVWNSNLGIQGRMSELSRKSRYONVWRPNSID-----SEYKALNVAFAFGK	172	
DB	131	TADILMGAIRNNRVAVAATKKG-SLVNTRFRFYFLIEDICPAATSTLEVNLQAVR----	184	
QY	173	RGIGSPAPSLSEKYHDSKDSAPAEVDLPGAVYGVKVCVFAPEPSGYLHITGAKAALLNK	232	
DB	185	-----EKKAAKAKEGASYSIALNTERKG-VVIRFPPEPSGYLHITGAKAALLND	232	
QY	233	YFA-ERYQRLIVRPDDTNPSKESNEFEENLKDITETGIKYDAVTTYSIDPYPKLMEAAE	291	
DB	233	YFAHEKYNGTLLVRFDDTNPSNEKLEFQALITEDIALMGIRKDKMSYSIDYDELYQVAL	292	
QY	292	SLIKGKAYIDTTPREOMRKEKMDIGESCRNNTBENSLMKEMVNGTERGMOCVGRK	351	
DB	293	QITIKGNVAADTTEKVAABQRMNGKPSGRDASVEENLAREEMKGGPBGIRMCIRAK	352	
QY	352	LDMODPNSLRDPVYVYRCNTDPHHRVGSKYKYPTYDFACPEVDALLEGYTHALRSSEYND	411	
DB	353	MSVDNPNNAMRPPVYIRCNRPAPHHHTGTWKAKYIPTYDFACPIVDSIEGYTHALRIERYND	412	
QY	412	RNAQYRRLIDMGRLRVEIYEFESRLNVYTLISKRKLMLFVONKVKYEDWTDPRFPVTQGI	471	
DB	413	RNPQYQWFLDTLKLHVOIWDPEARMMFIRITLLSKRKLTLVANGVGVWGDMDRPFRTIRGI	472	
QY	472	VRRGLKVELLIQPIIQGASKRLNLMEMWKLMTINKKIIDPCARHTAVALKQRYVFTLT	531	
DB	473	RRRGWITPILRFELIKQGPSKRIYNLDWTLIWAATNKYIDPAAPRTALILKKOMVAIVK	532	
QY	532	NGPEEPFVAILDRHHKFEAGAKKATTFARIMLDVADAANKGEEVTLMDGNMIVKEI	591	
DB	533	GGPAPFYTEERKRGKGNPAVGKKVYFGNTVIFPDQEDASFPQDEBITLMSGNAIVRKI	592	
QY	592	KVE--SGVITELVGLHIEGSAVTKTKITWLA-DIEELVPLSLVEFDVLISSKKLEDE	648	

Dd : |::||| ||||||| : |::||| : |::||| : |::||| :
593 ETDPSGIVKLELHLEHGEFKEKYTWLSTEGODILPELVDFVLANKOTIQEDD 652
Oy 649 DFLDNLNCTRRRIIPALGDANNRNIRKGIIQLERGGYRCDAPIRSSKPVLFAPIDG 708
Dd 653 VLEDVLNKTREPRADAVADCNVAEIKGEDDIQFERKKGYRVDRAYV-PEKPAVLFWIPTG 711
Oy 709 R 709
Dd 712 K 712
RESULT 11
ABU26317
ID ABU26317 standard; Protein; 715 AA.
AC ABU26317;
DE 16-APR-2003 (first entry)
XX Aspergillus fumigatus essential gene protein #975.
KW Fungicide; cyostatic; essential gene; Aspergillus fumigatus; infection;
OS cancer; contamination; biofilm; antibody; immune response.
PZ Aspergillus fumigatus.
MO WO200286090-A2.
FI 31-OCT-2002.
IN 23-APR-2002; 2002WO-US13142.
PR 23-APR-2001; 2001US-285697P.
PP 27-APR-2001; 2001US-287066P.
PP 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
(ELIT-) ELITRA PHARM INC.
Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
WI, 2003-093124/08.
New purified or isolated nucleic acids of essential genes of Aspergillus fumigatus, useful for treating or preventing infections by A. fumigatus, or for treating a non-infectious disease in a subject e.g. cancer -
Disclosure; Page -: 175pp; English.
The invention relates to novel purified or isolated nucleic acids of essential genes of Aspergillus fumigatus. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as A. fumigatus, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or inhibit formation of an object by A. fumigatus, or to prevent or inhibit formation on a surface of a biofilm comprising A. fumigatus. The polynucleotides are useful for expressing recombinant protein for characterisation, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of A. fumigatus to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively

CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organisms invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This sequence represents a protein of one of the essential genes
 CC of *Aspergillus fumigatus* of the invention.

XX Sequence 715 AA;

Query Match 39.2%; Score 1465.5; DB 24; Length 715;
 Best Local Similarity 42.9%; Pred. No. 9.9e-121;
 Matches 309; Conservative 128; Mismatches 235; Indels 49; Gaps 14;

```

QY 4 ALSFSKSPISIIICAKLVGLPTINHSLAAGAPLQF--ASGSEHGVNPIILYAR 61
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 26 SINEAPSPVISTVEDTAV-----LREGDAVAVQYAGSANGPIFGILNAVQELRK 76
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 62 GASIASLSGRND-IEFGHVEWLEVPF--LSGSEFENACIFVDGFLASRTFLVGHDT 118
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 77 --DEPFLNSKDEKLE---NEWLSQLEAFAPLDFKALDPELQRLDTHLLRSFVVGYALS 130
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 119 IADIIVNLSLGIQGRWESLRKSKYQNLVWFNSID-----SEYKALNENVAAPVKG 172
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 131 TADIALKAIKIRGNRYAVAAIKKG-SLVAVTRFVFLDLCPMWISTLEVLQAVR----- 184
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 173 RGIKSPAPSLKEKVDKSPAPEVDLPKAVGKVCVFAPEPSGYLHGHAKALINK 232
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 185 -----EKKAKEGASDYIALNTEKG-VVIRFPPEPSGYLHGHAKALINK 232
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 233 YFA-ERYQGLIVRFDITNPSKESNEFEVNLKDIETLGIKYDAVYTSYDPKIMEMAE 291
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 233 YFAHEKYGTLIVRFDITNPSKESNEFEVNLKDIETLGIKYDAVYTSYDPKIMEMAE 291
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 292 SLILOGKAYIDTDPKEOMRKERMOGIESCRNNTVEENLSIMKEVNTGEMGCCVRGK 351
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 293 QIIDGNAYADDTKEVNAEQPMNGKPSKRRDASVEENLAFEEKKGTPEGLMCLIRAK 352
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 352 LDMODPNKSLRDPVYRCNTDPHHRVGSRYKVPYDACPFDVDALEGVTHALSSSEYHD 411
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 353 MSVGNPKAMRDPVYRCNPPHRTGTGKMTIPIYDACPISVDSIEVTHALRTIERD 412
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 412 RNAQYRILQDMGLRVEIYEPSRLIMVYLLSKRKLMPYQNKKEVDWTDPRPFTVQGI 471
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 413 RNPQVWFLDTLKLRLHVQIMDFARNFIRTLSSKRLTKLVNQGVMGMDSPRFTIRGI 472
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 472 VRBGLKVALIQLTIQOGASKNLMEWDKMTINKIIDIYVCARHTAVLKDQRIPTLT 531
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 473 RRRGWTIPALREFFILKQSPSKNITLMDWTLLWATNKXYIDPAHPHTAILKKDWKALVK 532
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 532 NGPEEPFVRILPRHKKPFEGAGKATTPANRIMLDVADAALINKGEVTLMDGNATVKEI 591
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 533 GGPATPYIEEKPFGKGNPAVGMKKVYFGNTVIFDQEDAKSKRQDEBITLMSGNATVKEI 592
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 592 KVE--SGVITELVGLHLEGSVTKTKITWLA-DIEELVPLSLVEFDYLLSKKLEBDE 648
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 593 ETDPTSGIVKELHELHLEHEDPFKTEKKVYTWLSTEGQDLIPELVDFYLLNKDTLQDGD 652
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 649 DFLDNLNCTRRBELPALDAMNRNIKRGETITQERKGYRCDAPITRSKKPVLPALPDG 708
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 653 VLEVLNKNTEFRDADVADCNVALEKEDIIQFERKGYRVDRAYV-GKDAVLENIPTG 711
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 709 R 709
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 712 K 712
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 12
 AAU40163
 AAU40163 standard; Protein; 581 AA.

XX AAU40163;
 XX 13-FEB-2002 (first entry)
 XX

DE Propionibacterium acnes immunogenic protein #1059.
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAM, Peking DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59510.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -

PS Example 1; SEQ ID No 1358; 1069pp; English.

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 581 AA;

Query Match 18.7%; Score 697.5; DB 22; Length 581;
 Best Local Similarity 32.0%; Pred. No. 1.1e-52;
 Matches 186; Conservative 98; Mismatches 222; Indels 75; Gaps 17;

```

QY 158 YKEALNE--VVAAPVGRKGIKSPAPSLKEKHVSQPSAPENVLPKAVGKVCVFAPE 215
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 4 YRDLPLDSGKVPAAVAPRG--NPAESSDPITHOVVADIQQ-DTYG--GRVGRFPPE 56
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 216 PSGLYHGHAKAALINKYFAERYQGLIVRFDITNPSKESNEFEVNLKDIETLGIRYDA 275
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 57 PNGYLHGHAKAYITDGVADBEDFGTCNLRLDITNPTETREYVESITLADIEMLGYSYPAH 116
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 276 VTYTSYDYPKLMEMAESLIQKAYIDTDPKEOMRKER---MDGISRCRNNTVEENLS 331
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 117 VVHASDYFEQLYEWAKYLVREGLAAYDDQSPETIREQGGYKGISSPYRNRPAAESLN 176
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 332 LMKEMVNVGTERRGMCVCVRGLDMQDPKSLRDPVYRCNTPHHRVGSKKVYVTTYDFAC 391
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

Db 177 LRRMRAGEFPDGSRCRLARIDQAMENWMDPVYTRRHQAHSSTGTEKCIPTPYDMAH 236
Qy 392 PFVDALBEGVTHALSSSEYHDNNAQYRILODMGLRVEI-----YEFSLNMYTLLS 444
Db 237 GQSDAIEGVTHSLCSLEFNS-----HRPLVDMFLSHLPLDGPAPKQREFALELTHITTS 291
Qy 445 KRKLIMFQNKVQKVEDMTDPRPTVOGIVRGLKVEALIQFLQGCASKNLMMENDKMT 504
Db 292 KRLKSLVTNNIIVGMDPRMPTLGMRRRGYPAAAIRAFQCAVATTGNSVKALEEFES 351
Qy 505 INKKIIDVCARHTAVLKDQSVI-----FTLTNGPEPPRIILPRHKKF 548
Db 352 FVRRELNATQRRNAVHLPLKVLVDGMPTDDGDPVWFQLVNPNP----- 400
Qy 549 EGAGKATTFANRIWLDYADA-----AINKEEYTLMDGNAIYK--IKESGV 597
Db 401 -DDGTRRVPTFGELMIEADDFREDPPRKFPILSPGHEVRLR--GAYLVATADVYKNPDGT 457
Qy 598 ITELVEGLHLEGSVYKT---KLKIT--WLADIEELVPLSLVEPYLISKK-KLEBEDF 650
Db 458 IAEVNASYDPOSRGCTADGRKVKSTMTMWS-AGHAIVTANLYDRLSASRIPGSQTGEA 516
Qy 651 LDNLNPCRREI-PALGDANNENIKRGELIERKGYRCD 690
Db 517 LDDLNPHERETLVEVMAEPALANVAPGEVQPERLGYPAD 557

```

RESULT 13

ABU06096 standard; Protein; 562 AA.

ABU06096;

22-JAN-2003 (first entry)

N. meningitidis vaccine antigen #72.

Vaccine; antigen; meningococcal disease; pathogenic bacteria;

meningitis.

Neisseria meningitidis group B.

MO200277648-A2.

03-OCT-2002.

22-MAR-2002; 2002MO-GB01399.

22-MAR-2001; 2001GB-0007219.

(MICR-) MICROBIOLOGICAL RES AUTHORITY.

(IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

Robinson A, Goringe AR, Hudson MJ, Bracegirdle P, West DM;

Oliver KU, Kroll JS, Langford PR;

WPI; 2003-018958/01.

N-PSDB; ABX09930.

Identifying an antigen for manufacturing a vaccine against

meningococcal infection, comprises contacting antibodies with

polypeptides, detecting polypeptide-antibody complexes, and identifying

bound polypeptides as antigens -

Claim 37, Page 309-310; 310pp; English.

The invention relates to identifying an antigen comprising:

(a) obtaining antibodies against a commensal bacteria, or an extract

from a commensal bacteria; (b) contacting the antibodies with

polypeptides obtained from an expression library of either a commensal

or pathogenic bacteria; (c) determining whether the polypeptides bind to

identifying that polypeptide as an antigen. Also included are:

(1) a method of preparing a vaccine composition, comprising identifying an antigen with the above method, and combining the antigen with a carrier; (2) a vaccine composition obtained by the above methods; (3) an antigen identified by the above methods; (4) a polypeptide encoded by all or a part of a nucleic acid sequence comprising the Neisseria lactamica DNA sequences detailed in the specification; (5) a vector comprising the nucleic acid molecule; (6) a method of preparing a composition for vaccination against infection by pathogenic bacteria, comprising: (a) obtaining a first antigen from a commensal Neisseria; (b) comparing the amino acid sequence of the first antigen with the amino acid sequence of the second antigen from a pathogenic bacteria, or comparing the sequence of a nucleic acid which codes for the first antigen with the sequence of the nucleic acid that codes for the second antigen; and if the first antigen is homologous to the second antigen or if the nucleic acid sequence for the first antigen is homologous to that of the second antigen, and (c) preparing a composition for vaccination against bacterial infection comprising the first antigen; and (7) an antibody that binds to the polypeptide antigen.

The method is useful in screening commensal and pathogenic bacteria for previously unidentified vaccine antigens by identifying polypeptide antigens that bind to sera raised against commensal bacterial proteins. The polypeptide is useful as a vaccine antigen which may be used in the manufacture of a medicament for vaccination against meningococcal infection (e.g. meningitis). The present sequence represents an antigenic protein from the pathogenic bacteria N. meningitidis.

SQ Sequence 562 AA;

Query Match 17.8%; Score 666.5; DB 24; Length 562;

Best Local Similarity 34.7%; Pred. No. 66-50;

Matches 184; Conservative 84; Mismatches 217; Indels 45; Gaps 14;

Qy 197 EVDLPKAVGKVCVRFAPPEPSGVYHIGAKALINKYFAERYQRLIVRPDTPNSKESN 256

Db 18 BEDLESKHTAVQVRPFEPBGVLIHGAISICNPLGAVLYDGLCNRPDTPNBEKND 77

Qy 257 EIVENLKDIEITIGIKY-DAVYTSDYPPKIMENASLIIKQKAYIDDPKQARKER-- 313

Db 78 EYVNAIKEDVWELGFHMGEPBRFASNYFDQLYDVAVGLIKQKAYVDLTPREMRERYGT 137

Qy 314 --MDGISRCNNVVEENTLSMKEMVNGTERGMCQCRGKLDMDPNKSLADPYRRNT 371

Db 138 LTEAGKNSPYDRDRAVEENLDLFTPKNGEFPDGSKTJRLKIDMSGNIMMDPVYIRLR 197

Qy 372 DPHHRVSGKYVYPTYPFACPFVDALBEGVTHALSSSEYHDNNAQYRILODM--GLRRV 428

Db 198 AHHNNTGDKCITYMYDTHTCISDAIGITHSLCTLFPEARHPLYDCVLDNIPAPHATRP 257

Qy 429 EYEFSLNMYVYLLSKRKLMPYQNKVQKVEDMTDPRPTVOGIVRGLKVEALIQFLIQ 488

Db 258 RQYEFSLRELLYITTSRKLNQLVKEKVGSGMDPRMPTTISGMRRRGVTPGLRLFARFA 317

Qy 489 GASKNLMMENDKMTINKKIIDVCARHTAVLADQVIFLTN--GPEEPFRIILPRH 545

Db 318 GISSENIIVDMSVLEGALIRELENSAPRLMVLPLKV--TLTNFETRTGS--RRAAFH 373

Qy 546 KKEFGAGKATTFANRIWLDYADA-----AINKEEYTLMDGNAIYK--EIKVE 594

Db 374 PNHEMGREVPISQIYIIEADDFPAENPPKFKXILPGGEVRLRH--QYVAKCEVAVDE 431

Qy 595 SGVITTELVEGL-----HLEGSVYTKLKITWLADIEELVPLSLVEPYLISKKGE-- 645

Db 432 AGNVVELKCSIDHDTLGNPEG--RKVGVIHWVS-AEHAIEIKVRLDRLFTVERPDVAV 488

Qy 646 --EBEDFL--DNINPCTRREIPALGDANNENIKRGELIERKGYRCD 690

Db 489 RGEQGEYLPFTDPLNPESVKEITAYAEPAADLPASBSRQPERIGYFTD 538

RESULT 14

ABP77077

ID ABB77077 standard; Protein; 573 AA.
 XX
 AC ABB77077;
 XX
 DT 07-MAR-2003 (first entry)
 XX
 DE N. gonorrhoeae amino acid sequence SEQ ID 684.
 XX
 KM Antibacterial; infection; vaccine; gene therapy.
 XX
 OS Neisseria gonorrhoeae.
 XX
 PN W0200279243-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 12-FEB-2002; 2002WO-1B02069.
 XX
 PR 12-FEB-2001; 2001GB-0003424.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Fontana MR, Pizza M, Masignani V, Monaci E;
 XX
 DR WPI; 2003-058415/05.
 XX
 DR N-PSDB; AB238047.
 XX
 PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 PT medicament for treating or preventing N. gonorrhoeae infection
 PS Disclosure; Page 239; 815pp; English.
 XX
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABB76736-ABP81046 represent nucleic acid
 CC molecules of the invention.
 XX
 SQ Sequence 573 AA;
 XX
 Query Match 17.7%; Score 661.5; DB 24; Length 573;
 Best Local Similarity 34.5%; Pred. No. 1.7e-49;
 Matches 183; Conservative 83; Mismatches 219; Indels 45; Gaps 14;
 QY 197 EVDLPKAVGVKVPAPBPSPGYLHGHAKAALNKYFAERYQGRLLVRFDDTNPSEKSN 256
 DB 29 EDDLSGSHGVQTRPFRPNQYLIHGHAKSICLNFGLAYIYDGLCNLRFDDTNPSEKND 88
 QY 257 EFVENLKDIEITLGIKY-DAVYTSDFPKLMEASLIKQKAYIDDPKEQNRKE- 313
 DB 89 EYVNAIKEDVEWMLGFHWAGBPRFASDYPDRLYDAVGLIKQKAYVDLDPKEBREVYRG 148
 QY 314 --MGISRCNNVVEEVLSTMKEMVNTBERGMQCVKGLDMQDPKSLADPPYYRNT 371
 DB 149 LTERAGKSPYDRSIEENLDLFTRMKNGBPDSKTLMLKIDMAAGINMMDPVYIRRR 208
 QY 372 DPHRVGSKYKVPYTFYDFACFVDALLEGVTHALRSSEHDHNAQYRLLQD--MGLRRV 428
 DB 209 AHHNNTGDKMCIYPMYDTHGICSDAIBGITHSLCTLEBARPLDYDWLIDIPALHATRP 268
 QY 429 EIVFSSRLMVMYLLSKRKLMPVQNKVEDWTDPRFPTVOGIVARGKLVALEIQTILQ 488
 DB 269 RQYEFSSRLLELYTTSKRKLMLQVLEKIVSGMDPRMPTISGMRRRGVTPGVRLLFARA 328
 QY 489 GASKNLMEWDKMTINKKIIDPVCAGHTAVLQDQYIFLTN---GPEEPFARILPRH 545
 DB 329 GISSESEIVDMSVLEGALRELENSAPRLMVLNPLKY--TLTNFOAKRTGS--RRAAFH 384
 QY 546 KKFPGAGKATTPANRILWDYADAA-----AINKGEVTLMDMGNAIVK--EIKVE 594

DB 385 PNHEEMGDREVPVSQTIYIEADDFAEKNPPKGFKRLIDGGEVRLNH--GYVIKGEVVKDE 442
 QY 595 SGVITELVGEI-----HLEGSVKTKTKITWLDADIEELVPLSLVEFDYLSKCK--- 643
 DB 443 AGNVELKCSIDHDTLGKNPEG--RKVGVIHWVS-ABHAARIVRLYDLRFTERRGAV 499
 QY 644 LEEDEDFL---DNLNPTRRRIEIPALGDANMKRNGEIIQLERKGYRCD 690
 DB 500 RGEDGELPFTDFLNPESVKEITAVAPPAKADLPASRWQPERIGYFTVD 549
 RESULT 15
 ID AAB96155 standard; Protein; 577 AA.
 XX
 AC AAB96155;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Putative P. abyssi glutamyl- and glutaminyl-tRNA synthetase.
 XX
 KM Hyperthermophilic archaeon; hyperthermophilic protein.
 XX
 OS Pyrococcus abyssi.
 XX
 PN FR2792651-A1.
 XX
 PD 27-OCT-2000.
 XX
 PF 21-APR-1999; 99FR-0005034.
 XX
 PR 21-APR-1999; 99FR-0005034.
 XX
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PA (IFRE-) IFREMER INST FR RECH EXPL MER.
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
 PI Querellou J, Weissenbach J, Saurin W, Hellig R;
 DR WPI; 2001-126236/14.
 XX
 PT New nucleotide sequences isolated from Pyrococcus abyssi encode
 PT proteins useful in industry -
 XX
 PS Claim 7; Pages 793-794; 1657pp; French.
 XX
 CC The present invention relates to the genomic sequence of Pyrococcus
 CC abyssi (see AAF6431 and AHA41223-7) and P. abyssi proteins. P. abyssi is
 CC a hyperthermophilic archaeon, which is isolated from deep-sea
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.
 CC The proteins of the present invention have various potential industrial
 CC uses, since the proteins are stable at very high temperatures, some up to
 CC 110 degrees centigrade.
 CC Note: This patent is in the same patent family as W0200065062, which
 CC contains additional sequences as shown in AAB99132-AAB99143.
 CC AAT75503-AAH75920 and AAG66436.
 XX
 SQ Sequence 577 AA;
 XX
 Query Match 17.5%; Score 653; DB 22; Length 577;
 Best Local Similarity 30.8%; Pred. No. 9.9e-49;
 Matches 181; Conservative 92; Mismatches 252; Indels 62; Gaps 13;
 QY 131 IGQWESLRSKK---YQNLVWFNSID-SEYKALNEVVAFAVGRGIGKSPAPSLKE 185
 DB 38 IGENPELRSKAREIVPIINKVVEEVLSTLDEQAKLMEIV-----PEYFE 83
 QY 186 KVHDSKPSAPAEVDLPKAVGVKVPAPBPSPGYLHGHAKAALNKYFAERYQGRLLV 245
 DB 84 KKEEKKEKKGLPLPLPAKGVKVTFRAPNPDAFHLGNARAAALSYEYAAQMGKFLR 143
 QY 246 FDDTNPSEKSN--FVNLLKDIEITLGIKYDAVYTSDFPKLMEASLIKQKAYID 303

Db 144 FDDTDPKVKRPEPIFYEWIIEDEWLGIKPDEIYASDRLELYXKAEELIKMGKAVVCT 203
QY 304 TPKEQMKERMDGIESRCRNNTVEENLSLMEKMNVTGERGMOCCYRGKLDMDQPKSLRD 363
Db 204 CKPEKFRBLKDKGIPCPHDEPVEVQLERMRKMLNGEYKEGAVVRIKTDLNHPNPAVRD 263
QY 364 PVYRCNTDPNH-RVGSKYKYVPTYDFACPFVDALGVTHALRSSEYHNRNAQYRIIJD 422
Db 264 WPAIRIVDNPNHPRAKNRYVWPPLYNFASALDHGLVTHIFRQGEHLENETRQRYIYEX 323
QY 423 MGLRVEIYEFPSRLNMVYTLISKRLMFWQNKVYEDTDPFPFTVQGIVRGLKVEALI 482
Db 324 FGWEYPTVHHGRLSIEGVILSKSXTKRGIEGKYLGMDDPRLGITIRALRRRGILPEAIK 383
QY 483 QFIQOGASKRLNLMEMDKLMTINKKIIDPVCARTAVLKDQVIFTLTNGPEEPFVRL 542
Db 384 ELIEVGLKSDATVSWDNLAIRKLVDPANRYFFVA--DPVMEVEGAPE--FIKI 439
QY 543 PRHKKFEGAGKKATTF--ANRIWLDYADAALINKGEEVTLMDWGNALVKEIKVESGVITE 600
Db 440 PLHPDHPERGTRERLFTPGKEIYVSKDDLILKGSFVRLKDLFNVELIVEGK---IKA 496
QY 601 LVGELHLEGSVYKTKLKITWLAD--IBELVPLSLVEFDYLISSKXLEDEDDFLDNLNPG 657
Db 497 KFHSEFEYIARKNKMIMHWVEGRPCVILP---EGDELIVRKGLLEKD----- 543
QY 658 TRREIPALGDANMNIRKGEIIOLEKRYRCDAPIRSSKPVULFA 704
Db 544 -----ANVRKAGEIYQFERFGFVRIDK--IEGEKVVAIYA 575

Search completed: January 25, 2004, 17:09:46
Job time : 628 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 25, 2004, 16:58:20 ; Search time 71 Seconds
(without alignments)

426.088 Million cell updates/sec

Title: US-09-831-683B-10
Perfect score: 3734
Sequence: 1 MEALSFKSDSPISIIICA.....SSKEVLFAPIDGRQASLS 715

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents AA:*
2: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6C_COMB.pep:*
7: /cgn2_6/ptodata/1/1aa/6D_COMB.pep:*
8: /cgn2_6/ptodata/1/1aa/6E_COMB.pep:*
9: /cgn2_6/ptodata/1/1aa/6F_COMB.pep:*
10: /cgn2_6/ptodata/1/1aa/6G_COMB.pep:*
11: /cgn2_6/ptodata/1/1aa/6H_COMB.pep:*
12: /cgn2_6/ptodata/1/1aa/6I_COMB.pep:*
13: /cgn2_6/ptodata/1/1aa/6J_COMB.pep:*
14: /cgn2_6/ptodata/1/1aa/6K_COMB.pep:*
15: /cgn2_6/ptodata/1/1aa/6L_COMB.pep:*
16: /cgn2_6/ptodata/1/1aa/6M_COMB.pep:*
17: /cgn2_6/ptodata/1/1aa/6N_COMB.pep:*
18: /cgn2_6/ptodata/1/1aa/6O_COMB.pep:*
19: /cgn2_6/ptodata/1/1aa/6P_COMB.pep:*
20: /cgn2_6/ptodata/1/1aa/6Q_COMB.pep:*
21: /cgn2_6/ptodata/1/1aa/6R_COMB.pep:*
22: /cgn2_6/ptodata/1/1aa/6S_COMB.pep:*
23: /cgn2_6/ptodata/1/1aa/6T_COMB.pep:*
24: /cgn2_6/ptodata/1/1aa/6U_COMB.pep:*
25: /cgn2_6/ptodata/1/1aa/6V_COMB.pep:*
26: /cgn2_6/ptodata/1/1aa/6W_COMB.pep:*
27: /cgn2_6/ptodata/1/1aa/6X_COMB.pep:*
28: /cgn2_6/ptodata/1/1aa/6Y_COMB.pep:*
29: /cgn2_6/ptodata/1/1aa/6Z_COMB.pep:*
30: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
31: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
32: /cgn2_6/ptodata/1/1aa/6C_COMB.pep:*
33: /cgn2_6/ptodata/1/1aa/6D_COMB.pep:*
34: /cgn2_6/ptodata/1/1aa/6E_COMB.pep:*
35: /cgn2_6/ptodata/1/1aa/6F_COMB.pep:*
36: /cgn2_6/ptodata/1/1aa/6G_COMB.pep:*
37: /cgn2_6/ptodata/1/1aa/6H_COMB.pep:*
38: /cgn2_6/ptodata/1/1aa/6I_COMB.pep:*
39: /cgn2_6/ptodata/1/1aa/6J_COMB.pep:*
40: /cgn2_6/ptodata/1/1aa/6K_COMB.pep:*
41: /cgn2_6/ptodata/1/1aa/6L_COMB.pep:*
42: /cgn2_6/ptodata/1/1aa/6M_COMB.pep:*
43: /cgn2_6/ptodata/1/1aa/6N_COMB.pep:*
44: /cgn2_6/ptodata/1/1aa/6O_COMB.pep:*
45: /cgn2_6/ptodata/1/1aa/6P_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1506	40.3	1512	4	US-09-443-184-48
2	1437	38.5	1440	3	US-09-357-251-37
3	708.5	19.0	589	4	US-09-252-991A-30001
4	528	14.1	313	4	US-09-328-352-7010
5	212.5	5.7	516	4	US-09-198-452A-598
6	212	5.7	484	3	US-08-913-578-2
7	212	5.7	484	3	US-08-785-427-2
8	209	5.6	483	4	US-09-107-532A-5896
9	205.5	5.5	480	2	US-08-962-203-2
10	205.5	5.5	480	3	US-09-282-125A-2
11	205.5	5.5	480	4	US-09-273-142-2
12	203	5.4	501	4	US-09-134-001C-4115
13	199	5.3	348	2	US-08-844-153-2
14	193.5	5.2	287	4	US-09-328-352-7096
15	187	5.0	502	4	US-09-252-991A-28736
16	185.5	5.0	362	4	US-09-252-991A-26917
17	171	4.6	524	4	US-09-328-352-6089
18	133	3.6	1275	3	US-09-120-513-2
19	133	3.6	1275	3	US-09-450-105-2
20	132.5	3.5	313	4	US-09-328-352-5463
21	119.5	3.2	1627	4	US-09-328-352-6604
22	116	3.1	804	1	US-08-785-428-2
23	116	3.1	804	2	US-08-996-797-2
24	115.5	3.1	2662	4	US-09-595-684B-31
25	114	2.9	574	3	US-09-352-990-14
26	110	2.9	479	4	US-09-134-001C-4128
27	110	2.9	515	4	US-09-107-532A-5317

28	108.5	2.9	2172	1	US-08-611-107-31	Sequence 31, Appl
29	108	2.9	3959	2	US-08-970-269A-30	Sequence 30, Appl
30	108	2.9	3959	2	US-09-407-562-30	Sequence 30, Appl
31	106	2.8	466	1	US-08-785-066-2	Sequence 2, Appl
32	106	2.8	466	3	US-09-007-355-2	Sequence 2, Appl
33	106	2.8	466	3	US-08-913-489-2	Sequence 2, Appl
34	106	2.8	466	3	US-07-966-278-1	Sequence 1, Appl
35	106	2.8	775	1	US-08-424-921-1	Sequence 1, Appl
36	106	2.8	775	2	US-08-556-355A-1	Sequence 1, Appl
37	106	2.8	775	2	US-07-803-672A-1	Sequence 1, Appl
38	106	2.8	775	4	US-09-244-889A-1	Sequence 1, Appl
39	106	2.8	776	2	US-08-688-649-37	Sequence 37, Appl
40	106	2.8	2257	1	US-08-611-107-10	Sequence 10, Appl
41	106	2.8	2257	2	US-08-422-560A-10	Sequence 10, Appl
42	106	2.8	2257	3	US-08-468-793-10	Sequence 10, Appl
43	105.5	2.8	1683	3	US-08-755-587-183	Sequence 183, App
44	104	2.8	946	3	US-08-560-005-4	Sequence 4, Appl
45	104	2.8	946	3	US-09-418-540-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-443-184-48
Sequence 48, Application US/09443184A
Patent No. 6372431
GENERAL INFORMATION:
APPLICANT: Cunningham, Mary Jane
APPLICANT: Zweigler, Gary
APPLICANT: Kaser, Matthew R.
APPLICANT: Panzer, Scott
APPLICANT: Selhammer, Jeffrey J.
APPLICANT: Yue, Henry
APPLICANT: Baughn, Mariah
APPLICANT: Azimzai, Yalda
APPLICANT: Lal, Preeti
TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS
FILE REFERENCE: PC-0007 US
CURRENT APPLICATION NUMBER: US/09/443,184A
CURRENT FILING DATE: 1999-11-19
NUMBER OF SEQ ID NOS: 138
SOFTWARE: PERL Program
SEQ ID NO 48
LENGTH: 1512
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6372431 2302721CD1
US-09-443-184-48
Query Match
Best Local Similarity 42.2%; Pred. No. 1.1e-143;
Matches 305; Conservative 146; Mismatches 232; Indels 40; Gaps 9;
4 ALSSKSDSPISIIICAALVGLPTLINSGLASGAPLQFASGSLGCVNPIIYIARCA 63
5 SLTVNSGPPGALLAVHVDVSI--SVBEGENILHSENVIFTDVNSILRLAVVA 62
6 SLASISGNDIEFGHVWELVAPTLPSGSE--FNNACLFVQGLFASRFLVGHGLTINDI 122
7 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
8 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
9 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
10 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
11 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
12 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
13 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
14 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
15 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
16 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
17 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
18 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
19 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
20 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
21 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
22 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
23 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
24 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
25 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
26 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
27 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
28 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
29 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
30 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
31 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
32 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
33 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
34 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
35 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
36 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
37 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
38 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
39 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
40 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
41 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
42 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
43 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
44 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122
45 TTAGLYGSLMEHREIDHWFESATKSSCDSTFTINELHCHSLKTYLVGNSLSLDL 122

Dh	103	RVAPEKKDQVGFVEJLPGAMGKVTVRPEPPASGLYHGHAKALLNQHGVAFKGLIM	162
Qy	245	RPDDNTPSKESNEFEJENILKJOLETIGUYDAVYTTSDVFPPLMAMASGLKOGAYYDDT	304
Dh	163	RPDDNTPBEKEDFEFVEJLVEDAMAHYPRDPTTISDFETIMYAKELIOEGKAYVDDT	222
Qy	305	PKQMKERKEMDJESRCRNNTVEENL5LMKEMVNGTERGMOCVYRGKJLMDQPNKSLRP	364
Dh	223	PAQOMAEAREQJTESKHKRNPIEKVLQWMEEMKKGSGHSCCRAXIKIDMSNNGCWRDP	282
Qy	365	VYVRCTDHPHHVSGYKXKYPYTPDAPCFVDAJLBSGYTHALPSSSYHBNQYRYIIODMG	424
Dh	283	TLVRCKIOHPRTGNKXNYVPTPYDACPJVDISIGVTHALPTEYHORDQFWITIEALG	342
Qy	425	LRREVEJESRLNMYVTTLSKRLKLMFQONKKVEDWTDPREPTYOQJIVRGLKVEALIQF	484
Dh	343	IRKPEYIMESRLNLTNNYVLSKRLKLMFVNEGJVGMDPRPPTYRGLVRGMVTBGLKQF	402
Qy	485	ILQOGASKULNMEMDKLMTINKKIIDVCARHVALKQORVITTLNGBEPFVRLPR	544
Dh	403	IAAQGSSRSVVMEMDKTMAFNKKYIDVAPRYVALAKKEVI PUNVEAOE--MKXVAK	460
Qy	545	HKKFEGAGKATTPFANRILDYDAADAAJNKKSEBVTJLMDMGALIVEKI-KYESGYVTEJVG	603
Dh	461	HPKNPEVGLKPYWYSPKPIBGAJDLPTSBEEMTFLINMGMLNTKJHNKADGKJISLDA	520
Qy	604	ELHLEGSVYTKJLKTWLADIEELVPSL--VEPDYLSKKKLEEDDFDLNLPCTRRE	661
Dh	521	KFJLEKRDYKTKTKYTMALFETHALPIFIVICVTEHLLTKFVLKEDBDFQOYVKNKSKHE	580
Qy	662	IPALSGANNRNTKRGJIIOLBERGKYRCD-----APRIRSKSVLVAJFALPDG	708
Dh	581	ELMLGSPCKJLKDGDIIOLQKRGFFCDQYQYEPVSPISCEABCVLIYIPDG	633

RESULT 3
US-09-252-991A-30001
; Sequence 30001, Application US/09252991A

```

: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196, 136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ. ID NOS: 33142
: SEQ ID NO 30001
:
: LENGTH: 589
:
: TYPE: FRT
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-30001

```

[illegible]

QY 351 KLDMDPNKSLRDPVYRCNTDPPHHRVSGKYKVFPTYPACFVDALSGVTHALSSSYH 410
DB 210 KIDMOSPMNNLRDPIYLRIRHAIHHQTGDKMCIYPSYDFTHQSDAIEGITHSICITLFE 269
QY 411 DRNMQYIILQDMGL-RVEIYEFSRLNMVNTLSKRLKLMVQNKRYEDWTDPFPYVQ 469
DB 270 DHRPLWFLANLPPAPQPROYEFRRLNLTVTSTKRLKOLVDEGHSVGMDDPMSITLS 329
QY 470 GIVERGLKVEALIOFTLOOGASKNLMLMWDKLTMTINKKIIDPPVCARHTAVLKQORVFT 529
DB 330 GYRRRGYRPESIRNCEMIGVNRASGVVDIGMLSESIDHDATAPRMCAUKPLKVV-- 387
QY 530 LTNGEPEEPVRI-LPRHKKFEGAGKATTPANRIMWYADAANIN-----KGEVYTL 580
DB 388 ITNPEGQVENLELPRHPK-EDMGVAVLPFGRELPTDAGDFEVPVAGYKRLIPGGEVRL 446
QY 581 MDKGAIVK---EIVESGVITEL-----VGLHLEGSVKTKLKITWLADEELVP 629
DB 447 R-GSYVIRADEAIDADGNVIELRCSYDPTLGK-NPEG--RKVKGVYHNV-PAEGSVE 500
QY 630 LSLVFPDYLI---SKKLEDEDFLDNLNP-----CTREIPALGDAMNRNKRGDI 678
DB 501 CEVRLYDLRFSANPEKABEGSFLDNTNADSLQVLACG--RAEPLDQANPE-----DR 553
QY 679 IQLERKGYRCDAPFIRSSKPV 700
DB 554 FOFEREGYFVADLKDSRPGKP 575

RESULT 4

US-09-328-352-7010
; Sequence 7010, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC9-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7010
; LENGTH: 313
; TYPE: PR1
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7010

Query Match 14.1%; Score 528; DB 4; Length 313;
Best Local Similarity 37.5%; Pred. No. 4,2e-45;
Matches 119; Conservative 55; Mismatches 113; Indels 30; Gaps 7;

QY 163 NEVVAAPVKGKIGKSPASLSKEKYHDSKDPSPV-----DLPGAIVKVC 209
DB 6 NDVVS-----NLPPNPTPN-----THASVDSAQOQAGLDPVVRQVITDDLAAGRAKQIV 55
QY 210 VRPAPEPSGYLHIGAKKALLNKYFAERYQGRLLYRFPDTNPSKESNEFENLKDIEITL 269
DB 56 TRPPEPEPGYHIGHVKALICLNFVGAEEFNGLCNRPDPTNPDABEQQYVDGIANDVAKL 115
QY 270 GIKYDA-VTYTSDYFPKLMEMAESLIKQKAYIDTTPKORKEKMDIE-----SRCRN 324
DB 116 GFSNNGEPRVAGYFDOLYAWAIQILIEGDAYVDLQSPBEIKLNGSVEEGKNSPYDA 175
QY 325 TVEENLSIMKEMVNGTERGMQCCVRGKLDMDPNKSLRDPVYRCNTDPPHHRVSGKYK 384
DB 176 SVEENLSIMKEMVNGTERGMQCCVRGKLDMDPNKSLRDPVYRCNTDPPHHRVSGKYK 384
QY 385 PTYDFACFVDALSGVTHALSSSYHDMNAQYRILQDMGRRV-EIYEFSRLNMVNTLS 443
DB 236 PMYDIAHPLSALIBGITHSLCTLEFQDHRPFYDWIVKVSKAAPHOYESRLNVDYIT 295
QY 444 SKRKLWVQNKRYEDW 460

DB 296 SKRKLWVQNKRYEDW 460

RESULT 5

US-09-198-452A-598
; Sequence 598, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 598
; LENGTH: 516
; TYPE: PR1
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-598

Query Match 5.7%; Score 212.5; DB 4; Length 516;
Best Local Similarity 25.6%; Pred. No. 1.6e-12;
Matches 97; Conservative 65; Mismatches 146; Indels 71; Gaps 16;

QY 208 VCVRFAPEPSGYLHIGAKKALLNKYFAERYQGRLLYRFPDTNPSKESNEFENLKDIE 267
DB 17 VRVVAASPTGDPHVGTAYMALFNEIFAKPKGKMLIRIEDPTRRQDYENIFSLR 76
QY 268 TLGIKYDA-----VTYTSDFPKLME-MAESLIKQKAY-IDTTPKE--QMRKER-- 313
DB 77 WCGIQMBGPRVGPYGRYQSEERTKIYQGVETLTLTDCAKCFAPFQELABRAVA 136
QY 314 ---MDIESRCRNTVEENLSIMKEMVNGTERGMQCCVRGKLDMDPNKSLRDPVYRCN 370
DB 137 LGYRGYDRRYRYSPE-----EVASREMAQGPYIR-----LKVPLSGCV 178
QY 371 -----TDPHHRVSGKYKVPPTYDFACFVDALSGVTHALSSSYHDMNAQY 417
DB 179 FEDYSKRRVVPWADVDDVLVKSDDGPTYHFAVVIDHMLGITHVLRGESLSTPRL 238
QY 418 RILODMGLR-RVEIYEFSRLNMVNTLSKRR-----LLMPVQNKRYEDWTDPFPYVQ 472
DB 239 LLYAFGEPEPVFLMPLLNLPDGTKLSKKNPSTIYYDSGYKE-APVNFITLMGYS 297
QY 473 RRG---LKYBALIOFTLOOGASKNLMLMWDKLTMTINKKIID---PVCARHTAVLKQ 524
DB 298 MEGDEEYVSLERITETENPRIGSGAVFDIQKLDMMNKYILNHEGSPCC-----LLKEL 352
QY 525 RVITTLNGEPEEPVRI 543
DB 353 Q--GMLLN--DEFLKILP 367

RESULT 6

US-08-913-578-2
; Sequence 2, Application US/08913578
; Patent No. 6218159
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; TITLE OF INVENTION: No. 6218159el tRNA synthetase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913.578
FILING DATE: 17-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9601069.9
FILING DATE: 19-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31352
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 484 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-913-578-2

Query Match 5.7%; Score 212; DB 3; Length 484;
Best Local Similarity 20.8%; Pred. No. 1.6e-12;
Matches 106; Conservative 92; Mismatches 201; Indels 110; Gaps 18;
QY 207 KVCVFAPEBPGYHIGHAKAALLNKYFAERYQGLIVRPDDTNPSKESNEFVENLKDI 266
DB 4 RIRVRYAPSPGYLHIGNARTALFNVLAKYNGDFVIRIEDTKKNLEDEGETSQFDNL 63
QY 267 ETLGIXYD-AVTYTSDFPK-----LMEMASLTKOGKAYIDTTPKEQMRKMDGI 317
DB 64 KWLGLDWDSEVDKNGGYPYRQSERQHIYQPLIDQLAEDKAYCYWTEBLEAREKQI 123
QY 318 ESRGRNNTVEENLSLWKEVNGTE-RGMQCCVRGKLDMDPNKSLRDPVYRCNTDPHNR 376
DB 124 ARGEMPRYGGQHAHLEBQROQFEABGRQPSIRFRV-QNGTYSFDDWVKGNISFD-SNG 181
QY 377 VGS---KYKVPYYPDACPFDVALEGVTHALRSSEYHDNRNAQYRILQDMGLRREVIYE 432
DB 182 IGDWVIYKDGIPYTNFAVAIDHWMQISDVIRGDDHISNTPKQIMIEARGW--EPPR 238
QY 433 FSRLLMYT---LISKR--KLWLVQNKYVEDTDPFRFPYQGIIVRGKLVKVEALIQFIL 486
DB 239 FGHMSLVNERKKLSKRDGQILOF-----IEQYRD-----LGYLEALFNPIA 282
QY 487 QQGSKNLNLMEWDLMTINKKIIDPVCARHTAVLKDQRVIFTLTNGPEEPFVRILPRHK 546
DB 283 -----LIGWSP-----EGEEEIFS-----KEEFIKIFDEKR 308
QY 547 KPEGAKKATTPANRIMLDYADAAAIKNGEEVTLMDGNALVKEIKVESGVITELVGLH 606
DB 309 L-----SKSPAFFDQKQLAWNNQYMKQDTEVFO---LALPHLIKANLIPVPSBED 359
QY 607 LEGSVKTTKLTITWLADIIEVLVPLSVFVDYLSKKLEBEDDFLDNINPCTREIRIPLG 666
DB 360 LSWGRKLIALYOKENSVAGEIVPLSEMPF-----KEMPALG 395
QY 667 DANMENIRGEIILQERKGYRCDA--PF 693
DB 396 EBEQVINGEQVPELMTHTLFLKLEALBEP 424

RESULT 7
US-08-785-427-2
Sequence 2, Application US/08785427
Patent No. 6238900
GENERAL INFORMATION:

APPLICANT: Hodgson, John
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6238900e1 crna synthetase
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785.427
FILING DATE: 17-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9601069.9
FILING DATE: 19-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31352
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 484 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-785-427-2

Query Match 5.7%; Score 212; DB 3; Length 484;
Best Local Similarity 20.8%; Pred. No. 1.6e-12;
Matches 106; Conservative 92; Mismatches 201; Indels 110; Gaps 18;
QY 207 KVCVFAPEBPGYHIGHAKAALLNKYFAERYQGLIVRPDDTNPSKESNEFVENLKDI 266
DB 4 RIRVRYAPSPGYLHIGNARTALFNVLAKYNGDFVIRIEDTKKNLEDEGETSQFDNL 63
QY 267 ETLGIXYD-AVTYTSDFPK-----LMEMASLTKOGKAYIDTTPKEQMRKMDGI 317
DB 64 KWLGLDWDSEVDKNGGYPYRQSERQHIYQPLIDQLAEDKAYCYWTEBLEAREKQI 123
QY 318 ESRGRNNTVEENLSLWKEVNGTE-RGMQCCVRGKLDMDPNKSLRDPVYRCNTDPHNR 376
DB 124 ARGEMPRYGGQHAHLEBQROQFEABGRQPSIRFRV-QNGTYSFDDWVKGNISFD-SNG 181
QY 377 VGS---KYKVPYYPDACPFDVALEGVTHALRSSEYHDNRNAQYRILQDMGLRREVIYE 432
DB 182 IGDWVIYKDGIPYTNFAVAIDHWMQISDVIRGDDHISNTPKQIMIEARGW--EPPR 238
QY 433 FSRLLMYT---LISKR--KLWLVQNKYVEDTDPFRFPYQGIIVRGKLVKVEALIQFIL 486
DB 239 FGHMSLVNERKKLSKRDGQILOF-----IEQYRD-----LGYLEALFNPIA 282
QY 487 QQGSKNLNLMEWDLMTINKKIIDPVCARHTAVLKDQRVIFTLTNGPEEPFVRILPRHK 546
DB 283 -----LIGWSP-----EGEEEIFS-----KEEFIKIFDEKR 308
QY 547 KPEGAKKATTPANRIMLDYADAAAIKNGEEVTLMDGNALVKEIKVESGVITELVGLH 606
DB 309 L-----SKSPAFFDQKQLAWNNQYMKQDTEVFO---LALPHLIKANLIPVPSBED 359
QY 607 LEGSVKTTKLTITWLADIIEVLVPLSVFVDYLSKKLEBEDDFLDNINPCTREIRIPLG 666

Db 360 LSWGRKLTALYQKEMSYAGEIIVLSEMF-----KEMPALG 395
Qy 667 DANNENIRGEIIQERKGYRCDA--PF 693
Db 396 EEOQVINGEQVPELMTHLFSLKLEALPEP 424

RESULT 8

US-09-107-532A-5896

Sequence 5896, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESS: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneka

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 893-5007

TELEFAX: (781) 893-8277

INFORMATION FOR SEQ ID NO: 5896:

SEQUENCE CHARACTERISTICS:

LENGTH: 483 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...483

SEQUENCE DESCRIPTION: SEQ ID NO: 5896:

US-09-107-532A-5896

Query Match 5.6%; Score 209; DB 4; Length 483;
Best Local Similarity 24.6%; Pred. No. 3.2e-12;
Matches 78; Conservative 45; Mismatches 100; Indels 94; Gaps 13;

Qy 210 VRAPESGYLHGHAKALLINKYFAERYQRLIVRFDTPNPSKESNEFVENLKDIETL 269
Db 3 VRAPSPGTHLIGNATFALFNLYLFAHNDGDFIIRIDTQKRNIIDGKSOLENLAWL 62

Qy 270 GIKXD-ATYTSQVFP-----KLEMAESLIKO-----GKAYIDDPKQOMKER----- 313
Db 63 GMMNDESENPGEYGYRQSRKEIYQPLIDQLASNAKYCYCTEELERERQARQ 122

Qy 314 -MDGIESCRNNTVENISLWKE-----MVNGTERGMQCCVAGKLDMDPNKSL 361

Db 123 EMPHYAGTCA/LTPEEQAE--KEAQGLSEVVRFRPRNTEYAFITDMVKALISFSDN--- 177
Qy 362 RDPVYVNCNDPHHRVGSKKV-----YPTDFACPFVDALLEGVTHALRSSEYHDRAQY 416
Db 178 -----IGDFVIQKRDGMPTYNFAVVDHLMKITHVLRGDD-HIANTP- 220
Qy 417 YRIQDMGLRVEIYEFSSRLNMVYTLSSKRLMFWONKKVEDTDRPFPTVQGIV--RR 474
Db 221 -----KQLMAYE-----AFGWTPTPTFGMTLINSET 247
Qy 475 GLKV---EALIQITLQ 487
Db 248 GKLSKRDESILQPIEQ 264

RESULT 9

US-08-962-203-2

Sequence 2, Application US/08962203

Patent No. 5976840

GENERAL INFORMATION:

APPLICANT: Jaworski, Deborah

APPLICANT: Lawlor, Elizabeth

APPLICANT: Wang, Min

TITLE OF INVENTION: NOVEL STREPTOCOCCAL ERS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESS: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/962,203

FILING DATE: 31-OCT-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/844,153

FILING DATE: 18-APR-1997

APPLICATION NUMBER: 9607992.6

FILING DATE: 18-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimmi, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P31459-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 480 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-962-203-2

Query Match 5.5%; Score 205.5; DB 2; Length 480;
Best Local Similarity 19.1%; Pred. No. 7.1e-12;
Matches 99; Conservative 82; Mismatches 170; Indels 167; Gaps 19;

Qy 210 VRAPESGYLHGHAKALLINKYFAERYQRLIVRFDTPNPSKESNEFVENLKDIETL 269
Db 1 VRAPSPGTHLIGNATFALFNLYLFAHNDGDFIIRIDTQKRNIIDGKSOLENLAWL 60

Qy 270 GIKYDAVTYT-----SDYFPKLEMAESLIKOGKAYIDDPKQOMKER----- 313

Db 61 GMDDESPESHENYRQSERLDLYQKTIQDLAEGKAYKSYTTEBELAERERQEAETP 120
Qy 314 -----MDGIESRCNNNTVEENLSMKEMVNGTERGMOCCVGRK 351
Db 121 RYINEYLGMSEBEKAAYIAEREAAGIIPVRLAVNESGIYKWHDM-----VKGD 169
Qy 352 LDMQDPNKSLLDPVYYRCNTDPHHRVGSKYV-----YPTDFACFPVDALGVTALRS 406
Db 170 IEFGGN-----IGGDWVIOKKDGYPTYNFAVVIDDHMOQSHVIRG 211
Qy 407 SEYHDRAQYRILQDMGLRVEIYEFSLNMYT-----LLSKR--KLWFOVKKVED 459
Db 212 DDHANTPKQMLMYEALGW---EAPFEGHMTLIINSETKLSKRDNTTLOFIEDYR--- 265
Qy 460 WTDPRFPVQGIVRKGLVEALIOFILOQASKNLIMEMDKLWTINKKIIDPVCARHTA 519
Db 266 -----KKGYLEAVFNFI-----LLGW-----NCGEDEIFSR--- 294
Qy 520 VLKQORVIFLTNGPEEPFVRILPRHKKFEGAGKATTPANRIMWDYADAAINKGEVY 579
Db 295 -----BELIKLPDENR-----LSKSPAAFQDK-KLDWMSNDYIKNADLET 333
Qy 580 LMDGNAIVKEIKVESGIVTELVEGL-----HLEGSVKTKLKITWLDIEEL----- 627
Db 334 IFE---MAKPFLEAGRLTDKAEKLPDYKPKMSVDEIIPLDLPFSDFPELTAEARE 389
Qy 628 -----VPLSLVEFDYLLSKKLE---EDEDPLDNLP 656
Db 390 VMTGETVPTVLEAF-----KAKLEAMTDEFEVTENIFF 422

RESULT 10

US-09-282-125A-2
; Sequence 2, Application US/09282125A
; Patent No. 6165760
GENERAL INFORMATION:
APPLICANT: Jaworski, Deborah
Lawlor, Elizabeth
Mang, Min
TITLE OF INVENTION: NOVEL STREPTOCOCCAL ERS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/282,125A
FILING DATE: 07-Aug-2000
CLASSIFICATION: <Unknown>
18-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/962,203
FILING DATE: <Unknown>
APPLICATION NUMBER: 9607992.6
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31459-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX: <Unknown>
18-APR-1996
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-282-125A-2

Query Match 5.5%; Score 205.5; DB 3; Length 480;
Best Local Similarity 19.1%; Pred. No. 7.1e-12;
Matches 99; Conservative 82; Mismatches 170; Indels 167; Gaps 19;

Qy 210 VRFAPEPSGYHICHAQALINKYFAERYQRLIVREDTPNPKSESNEPVENLLKDIETL 269
Db 1 VRVAPSPGLHIGNARTALFNLYAHNGGTFLIRIEDTRKHVDEGERSQLENRLWL 60
Qy 270 GIKYDAVYTT-----SDYFPKLMEMASLIKQKAYIDTPKQMRER----- 313
Db 61 GMDDESPESHENYRQSERLDLYQKTIQDLAEGKAYKSYTTEBELAERERQEAETP 120
Qy 314 -----MDGIESRCNNNTVEENLSMKEMVNGTERGMOCCVGRK 351
Db 121 RYINEYLGMSEBEKAAYIAEREAAGIIPVRLAVNESGIYKWHDM-----VKGD 169
Qy 352 LDMQDPNKSLLDPVYYRCNTDPHHRVGSKYV-----YPTDFACFPVDALGVTALRS 406
Db 170 IEFEGN-----IGGDWVIOKKDGYPTYNFAVVIDDHMOQSHVIRG 211
Qy 407 SEYHDRAQYRILQDMGLRVEIYEFSLNMYT-----LLSKR--KLWFOVKKVED 459
Db 212 DDHANTPKQMLMYEALGW---EAPFEGHMTLIINSETKLSKRDNTTLOFIEDYR--- 265
Qy 460 WTDPRFPVQGIVRKGLVEALIOFILOQASKNLIMEMDKLWTINKKIIDPVCARHTA 519
Db 266 -----KKGYLEAVFNFI-----LLGW-----NCGEDEIFSR--- 294
Qy 520 VLKQORVIFLTNGPEEPFVRILPRHKKFEGAGKATTPANRIMWDYADAAINKGEVY 579
Db 295 -----BELIKLPDENR-----LSKSPAAFQDK-KLDWMSNDYIKNADLET 333
Qy 580 LMDGNAIVKEIKVESGIVTELVEGL-----HLEGSVKTKLKITWLDIEEL----- 627
Db 334 IFE---MAKPFLEAGRLTDKAEKLPDYKPKMSVDEIIPLDLPFSDFPELTAEARE 389
Qy 628 -----VPLSLVEFDYLLSKKLE---EDEDPLDNLP 656
Db 390 VMTGETVPTVLEAF-----KAKLEAMTDEFEVTENIFF 422

RESULT 11

US-09-273-142-2
; Sequence 2, Application US/09273142
; Patent No. 6300119
GENERAL INFORMATION:
APPLICANT: Jaworski, Deborah
Lawlor, Elizabeth
Mang, Min
TITLE OF INVENTION: NOVEL STREPTOCOCCAL ERS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:

	TITLE OF INVENTION:	EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
	FILE REFERENCE:	GTC-007
	CURRENT APPLICATION NUMBER:	US/09/134.001C
	CURRENT FILING DATE:	1998-08-13
	PRIOR APPLICATION NUMBER:	US 60/064,964
	PRIOR FILING DATE:	1997-11-08
	PRIOR APPLICATION NUMBER:	US 60/055,779
	PRIOR FILING DATE:	1997-08-14
	NUMBER OF SEQ ID NOS:	5674
	SEQ ID NO	4115
	LENGTH:	501
	TYPE:	PRF
	ORGANISM:	Staphylococcus epidermidis
	US-09-134-001C-4115	
Query Match	Similarity	5.4%; Score 203; DB 4; Length 501;
Best Local	Similarity	20.0%; Pred. No. 1,4e-11;
Matches	Conservative	86; Mismatches 183; Indels 118; Gaps 18
QY	202	GAKV--GKVCRRPSPSGYHGHAKAALNKFAEYOGRLYRFPDTPNSKSNEV 259
		: : : : : : : : : : : : : :
Db	14	GVKIMSRIRRVAPSPGYLHINARVALNYLFKAHNGPFVRIRIDTDSKNLEDE 73
QY	260	ENLKDIETTGIXKD-VVTYSDFP-KLMEMAS-----SLIQGKAYIDDTPEQR 310
		: : : : : : : : : : : : : : :
Db	74	SSQPNDNLKLWLDWDESVDKDGFGPYROSERAEIYNPLIOQLLEDRAKYCYMTEBELE 133
QY	311	KER-----MDGISRCRNNTVEENLSLWKEMVNGTER 342
		: : : : : : : : : : : : :
Db	134	AERNAQIARGEMRPYGGOHMLTBEORQYEABGRKPSIRRVPRDQYTTFVDM----- 187
QY	343	GMOCVBRGLDMODPNKSLRDVPYYRCNTDPHRVG---KKYVPLYDFACPFVDAL 398
		: : : : : : : : : : : : : : :
Db	188	--VGKEISFEESDN-----IGDWVIYKDGVPPTNFVAVDHYM 224
QY	399	GUTHALLSSEYHDNAQYRILQDMGLRVELYERSRLMYTT----LLSKR-KLIMPV 452
		: : : : : : : : : : : : : :
Db	225	QISVIVIGDDHVSNTPKOLMYEAFGW--BAPRGHNSLIYNEERKGLSKRDGOILOPI 281
QY	453	QNKEVEWMTPD--RFPTVOGIVVRG---LKVEALIQFIQQGASKYLNMEDWLMTI 505
		: : : : : : : : : : : : : :
Db	282	EQYRDDLGVLPEALFNFTILGMSPEGEBEITSKEEFIKIPBKRLSKSPAMFDROKLAV 341
QY	506	NKKIIDPVCARHTVALKDQVIFTLTNG-----PEEPVRLIPRKKEGAGKAKAT 557
		: : : : : : : : : : : : : :
Db	342	NNQVMK-----TKDETFFELALPHLIKAMLIPENPSEKOREMGKILALYOKEMS 392
QY	558	FANRI---WLDVADAALKGEVITLMDGNALYKLVKESGVTVTELVEHLHGSAVKT 613
		: : : : : : : : : : : : : :
Db	393	YAGEIVPLESEMFEHEMPELGKDBOEVLO--GEQVPE-----LMMHLGYKLESIESFEA 443
QY	614	TKLK 617
		: :
Db	444	TEIK 447
		:
RESULT 13		
US-08-844-153-2		
	Sequence 2,	Application US/08844153
	Patent No.	5958734
	GENERAL INFORMATION:	
	APPLICANT:	Lawlor, Elizabeth
	TITLE OF INVENTION:	No. 5958734el Compounds
	NUMBER OF SEQUENCES:	6
	CORRESPONDENCE ADDRESS:	
	ADDRESSEE:	SmltnKline Beecham Corporation
	STREET:	709 Swedeland Road
	CITY:	King of Prussia
	STATE:	PA
	COUNTRY:	USA
	ZIP:	19046
	COMPUTER READABLE FORM:	
	MEDIUM TYPE:	Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,153
FILING DATE: 18-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607992.6
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE NUMBER: F31459
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-844-153-2

Query Match 5.3%; Score 199; DB 2; Length 348;
Best Local Similarity 20.3%; Pred. No. 1,9e-11;
Matches 67; Conservative 56; Mismatches 109; Indels 98; Gaps 9;

QY 210 VRAPESGYLHIGAKALLNKYFAERYQGLIVRPDPTNKSNEFVENLKDIEFL 269
DB 1 VRAPSPGLHIGAKATLENYLVAHHGFTFLIRIDTRKRAVEDGERSQLENLRL 60
QY 270 GIKDAVYV-----SDYFRLMAMASLTKQRAYIDTPKQMRER----- 313
DB 61 GMDWSESHENYKQSERLDLYQYKIDQLAEGAYSYVEEELAEEREOAAGETP 120
QY 314 -----MDGIESRCNNVTENISLWKEVNGTERGMCCVAGK 351
DB 121 RYNEYLQMSSEKRAYIAERAGIIPVRLAVNESGIYKKNH-----VKGD 169
QY 352 LDMODPNKSLDPVYRCNTDPHNRVSKYV-----YPTDFACFPVDALEGVTHALS 406
DB 170 IEFEGN-----IGGDWVYQKDGYPYFAVVIDHDMQISHVIRG 211
QY 407 SEYHNRNAYRILODMGLRVEIYEFRLNNVT-----LISKR--KLWVVOXKYED 459
DB 212 DDHANTPKQLMVAALGM--EAEFGHMTLIINSETGKKLSKXDVTWLOFIEDYR--- 265
QY 460 WTDPRFTVQGIVRGLKEALIOFILOQ 489
DB 266 -----KKGYLPEAVNFIALLG 282

RESULT 14
US-09-328-352-7096
Sequence 7096, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7096
LENGTH: 287
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7096

Query Match 5.2%; Score 193.5; DB 4; Length 287;
Best Local Similarity 27.7%; Pred. No. 4.8e-11;
Matches 78; Conservative 49; Mismatches 124; Indels 31; Gaps 12;

QY 435 RLNNVYTLISRKLLWFQNKVEDWTDPRPPTVQGIVRGLKEALIOFILOQASKNL 494
DB 5 RLNNVY-----VNW-LKGYVNGMDPMPVTSNRKRGFTREGLRDFCKRQVSKTD 56
QY 495 NMEWDKLTINKKIIDPVCARHTAVLKQDRIFTLTNGPEEPFVRLPRHKEGAGKK 554
DB 57 GIVVAVMLEPCTIRQSLNTARQAVLNLKV--TLTNLPED-LDLTHARHPND-MGER 112
QY 555 ATTPANRIWDYADAAL-NKGEVTLMDGNALVKE--IKVESGYITELVG-LHLEG 609
DB 113 VIPLTKELIYIDRKDFEEVPPKGFRLIPD-GEVRLRAYVYKDE-VIKDANGVIELKC 170
QY 610 SV-----KTKLKITWLDIEVLPLSVVERDYLSKKKEDEDFLNLNPT 658
DB 171 SIDPDTGKQPEGRKVGIVHWVSATKG-IPAEVRIIDRLFTEDPDTGDFLNLNPD 229
QY 659 REIIPALGDANMRNIKGEIITLERKGYRCDAPEFIRSSKEV 700
DB 230 LKVVQAVIEPRLAQAKEDRQFERBEGYFVADYDHTPEKV 271

RESULT 15
US-09-252-991A-28736
Sequence 28736, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196,136
CURRENT APPLICATION NUMBER: US/09/252,991A
FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 28736
LENGTH: 502
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28736

Query Match 5.0%; Score 187; DB 4; Length 502;
Best Local Similarity 22.1%; Pred. No. 6e-10;
Matches 92; Conservative 66; Mismatches 156; Indels 102; Gaps 18;

QY 200 LP-GAKVGVKCV--RPAPEPGYLHIGAKALLNKYFAERYQGLIVRPDPTNKSNE 256
DB 1 LPLGRVSMVTVRIRIAPSPGTGDHVGATAYIALNLCPARHGQGFILRIDTQLRSTR 60
QY 257 EFVENLKDIEFLGIKD-----AVTYSDYFPRLMAMASLTKQRAY----- 300
DB 61 ESEGGIDALMLWLGIDWDEGDPVGPYQSGRHLYKKYSDLEVEKHAFCFTPE 120
QY 301 -IDTPPEQM-RKE--RMDGIESRCNNVTENISLWKEVNGTERGMQ--CCV-----RG 350
DB 121 RLDAVRAEQMARKETPRDGHCHLPRDQVQRLAAGESHVTRKVPTEGVCVVPMLRG 180
QY 351 -----KLMDQDNKSLRDPVYRCNTDPHNRVSKY-----KVYPTVD 388
DB 181 DVEIPMDKMDQVLMKADGLPTFLAVVDHLMGITHVLGEMLPAPLILYEFYG 240
QY 389 FACP---FVDALGVTHALRSSEYHNRNAYR-----ILDMG----- 424
DB 241 WEQQLCYMPLLRNPKSKSKRNKPSITFERMGVLPQALLNVLGMSGMPDEREK 300
QY 425 -LRREIYEFPSRLNNVYTLISRKLLW-----FVQNKKVED-----WT-DPRF----- 465

Db 301 TLAEHIEHFDLSRVSLGGPIFDLEKLSWLNQGMIREQSVBEPAREVOKMALNPEYLMKIA 360
Qy 466 PTVGGIVRRGLKVEALIOFILOQGASKUNLMEWDKLTMTINKKIDPVCARHTAVL 521
Db 361 PHVQGRVENFSCIAIPLAGFFPSGSGVPLDASLFEHKKL-----DPTQVRQVLOL 408

Search completed: January 25, 2004, 17:15:58
Job time : 72 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 17:09:55 ; Search time 247 Seconds

(without alignments)
598.448 Million cell updates/sec

Title: US-09-831-683B-10

Perfect score: 3734

Sequence: 1 MEALSRKSPISIIICAA.....SSKEVLFAPDRQASLS 715

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 777136 seqs, 20673638 residues

Total number of hits satisfying chosen parameters: 777136

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1574.5	42.2	716	US-10-369-493-2175	Sequence 2175, Ap
2	1562	41.8	724	US-10-369-493-21897	Sequence 21897, A
3	1537	41.2	725	US-10-032-585-7526	Sequence 7526, Ap
4	1471.5	39.4	715	US-10-128-714-3375	Sequence 3375, Ap
5	1465.5	39.2	715	US-10-128-714-8375	Sequence 8375, Ap
6	1459.5	39.1	606	US-10-369-493-13278	Sequence 13278, A
7	1349	36.1	618	US-10-369-493-42210	Sequence 4230, Ap
8	1339	35.9	1149	US-10-369-493-52131	Sequence 5211, Ap
9	1287.5	34.5	717	US-10-104-047-3100	Sequence 3100, Ap
10	753	20.2	554	US-10-369-493-751	Sequence 751, App
11	710.5	19.0	537	US-10-369-493-9677	Sequence 9677, App
12	707	18.9	554	US-10-369-493-21266	Sequence 21266, A
13	704.5	18.9	554	US-10-369-493-13759	Sequence 13759, A
14	703	18.8	554	US-10-369-493-412	Sequence 412, App
15	684.5	18.3	573	US-10-369-493-9393	Sequence 9393, Ap

16	683.5	18.3	580	US-10-369-493-17582	Sequence 17582, A
17	674.5	18.0	573	US-10-369-493-9215	Sequence 9215, Ap
18	672.5	18.1	551	US-10-369-493-19380	Sequence 19380, A
19	672.5	18.0	569	US-10-369-493-8623	Sequence 8623, Ap
20	662.5	17.7	574	US-10-078-770-166	Sequence 166, App
21	658	17.6	556	US-10-369-493-7538	Sequence 7538, App
22	658	17.6	573	US-10-369-493-4778	Sequence 4778, Ap
23	653	17.5	571	US-10-369-493-21578	Sequence 21578, A
24	649	17.4	542	US-10-369-493-20659	Sequence 20659, A
25	624	16.7	570	US-10-369-493-1319	Sequence 1319, Ap
26	624	16.7	570	US-10-369-493-20384	Sequence 20384, A
27	589.5	15.8	553	US-10-369-493-1067	Sequence 1067, Ap
28	588	15.7	799	US-10-032-585-7457	Sequence 7457, Ap
29	573	15.3	786	US-10-369-493-6228	Sequence 6228, Ap
30	572	15.3	811	US-10-369-493-32597	Sequence 32597, A
31	561.5	15.0	775	US-10-153-668-300	Sequence 300, App
32	559	15.0	551	US-10-369-493-930	Sequence 930, App
33	537.5	14.4	553	US-10-369-493-1102	Sequence 1102, App
34	518.5	13.9	562	US-10-369-493-11272	Sequence 11272, A
35	480.5	12.9	550	US-10-369-493-18217	Sequence 18217, A
36	477.5	12.8	544	US-10-369-493-22958	Sequence 22958, A
37	464	12.4	540	US-10-369-493-13582	Sequence 13582, A
38	437	11.7	541	US-10-369-493-10580	Sequence 10580, A
39	424	11.4	433	US-10-078-770-164	Sequence 164, App
40	381.5	10.2	586	US-10-369-493-18607	Sequence 18607, A
41	364	9.7	383	US-10-078-770-170	Sequence 170, App
42	271	7.3	473	US-10-369-493-92	Sequence 92, App
43	255	6.8	469	US-10-369-493-3022	Sequence 3022, App
44	247	6.6	487	US-10-369-493-3071	Sequence 3071, Ap
45	243	6.5	463	US-09-815-242-11325	Sequence 11325, A

ALIGNMENTS

RESULT 1
US-10-369-493-2175
Sequence 2175, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 2175
LENGTH: 716
TYPE: PRT
ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2175

Query Match 42.2% Score 1574.5; DB 12; Length 716;
Best Local Similarity 49.5% Pred. No. 3.8e-139;
Matches 319; Conservative 102; Mismatches 201; Indels 23; Gaps 10;
QY 70 GKNDIEFGHVWLEIAPPTLSGS---EFENACLFVDFGLASTFLVHGTLTADIAVW 125
DB 81 GENDRSL--VESWETA-SALAGHNFLELSLAQDDHILMRSLFVGYSLTADSFSIW 137
QY 126 SNLAGIQRPESLRKSKYQNVKRWFPNSIDSEYEAALNEVAAVFGRGIGKSPAPLKE 185
DB 138 GALKSNMAAGAVR-TQOYFLARWYKFDQ--NAVSVTMEETKAVNISK-----KQ 188
QY 186 KVDHSKPSAPEVDLFGAKYGVKVCVRPAPPSGLTILGHAKAALINKYFAERYGRLIVR 245

```

Db 189 K---SSGPNY-EIGLPDAIDGKVYTRPPEBSGYLHIGAKAALLNQYFANKHGXILVR 244
Qy 246 FDDTNPSEKSENEFEVNLKDIETLGIKYDAVYTSDFPKLMEASLIGKQKAYIDTP 305
Db 245 FDDTNPSEKSENEFODALIEDVALLGKEDVYTSIDYITLHQYCVDMIKSGQAYADTD 304
Qy 306 KEQARKRMQOIESRCNNVTVEENLSLMKENVNGTERGMOCCVGLKMDOPNKSILRDPV 365
Db 305 VETHRHRTGEPKSHRDRPTEESLEIISEMDKSDVGLKNCIRAKTSYENPNAMRDPV 364
Qy 366 YRRCNTDPHHRVGSKYKVPYDFACPFVDALIEGTHALRSSEYHNRNAQYRLLIDMDGL 425
Db 365 IYRNLDPHHRGTGKRYAPYDFACPIYDLEGTALRTTEYRDRNPLQOMMIKANL 424
Qy 426 RVEIYEPSRLNMYTLLSKRLLMFVONKKYEDWTDPRPTVOGIVRGLKVEALLQPI 485
Db 425 RKIHWFEFSRNFRTLLSKRKLTEIVDHGLVWGDDPRFTVGRVRRRGMTIEALQOYI 484
Qy 486 LQOGASKNLNMEMDKLMTINKKIIDPYCARTAVLKNQRYIFTLTNGPEEPFVRLPRH 545
Db 485 VSQSPKNILTLMDTSPWATNKKIIDPAPRHTAVESGDVVKATIVNGPAAPYADRRPRH 544
Qy 546 KKEFGAGKATTFANRILWLDYADAANKGSEVTLMDGNALVEIKYE-SGVITELVGE 604
Db 545 KKNPELGKKSIFANEILIEGADAQSPKQDEEVLMDGNAYREINRDAQKTSKLJE 604
Qy 605 LHLEGSVTKTKLITWLADIELVPLSIVEFDYILSKKLEBEDFDLNDNPTREIRIPA 664
Db 605 LHLGDPEKTEKTYLWLTADTEDKTPVDLVDYILTKOKLEGEYKDFLTPQTEFHSFV 664
Qy 665 LGDANMNIRKGEIIOLEKGYRCDAPFIRSSKPVULFAIPDGR 709
Db 665 PADGIKNKLKGDIIQVERKGYIVDVPF--DGTQAVLNIPIDGK 707

RESULT 2
US-10-369-493-21897
; Sequence 21897, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21897
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-21897

Query Match 41.8%; Score 1562; DB 12; Length 724;
Best Local Similarity 45.7%; Pred. No. 5,9e+138; Indels 56; Gaps 14;
Matches 328; Conservative 121; Mismatches 212;

```

```

Db 135 AADVACMGALRSNGM-CGSIHKONKVDVNVSRWYTLLEMDPIFGEAHDFLSKSLLELKSA 193
Qy 174 GIGSPAPSLKEKHYHDSKDSAPREVDPGAKVGVYCRFAPBPSGYLHIGAKAALLNKY 233
Db 194 NVGR-----KKEHKNANF-----EIDLPADAKMGVYVTRPPEBSGYLHIGAKAALLNQY 243
Qy 234 FAERYOGRILYRFPDTPSKSENEFEVNLKDIETLGIKYDAVYTSDFPKLMEASL 293
Db 244 PAQYVKGSLIIRFPDTPSKSENEFEVNLKDIETLGIKYDAVYTSDFPKLMEASL 303
Qy 294 IKQKAYIDTPPKQOMKERMDGIESRCNNVTVEENLSLW-KENVNGTERGMOCCVGRKL 352
Db 304 IKQKAYCDDTPPKQOMKERMDGIVASARORSVEENIRIFPEKMGNGTEBGLKQCVRAKI 363
Qy 353 DMQPNKSLRDPVYRRCNTDPHHRVGSKYKVPYDFACPFVDALIEGTHALRSSEYHNR 412
Db 364 DYKALNLTLRDPVYRRCNLTPHHRGTSTWKIYPTPYDFCVPIVDAIEGTHALRTIEYDR 423
Qy 413 NAOYRRLIDMGJARVEIYEFSRLNMYTLLSKRLLMFVONKKYEDWTDPRPTVOGIV 472
Db 424 NAOYDMMLQALRLKRVHIMDFARINFRTLLSKRKLQMMVDKLVGNMDDPRFTVGRVR 483
Qy 473 RGLKVEALLQPIIQOGASKNLNMEMDKLMTINKKIIDPYCARTAVLKNQRYIFTLTN 532
Db 484 RRGMTVEGLRNFVLSQSPSRVINLWNNLWAFNKKVYIDPAPRHTAVLNVFVKIHLBGE 543
Qy 533 GPEEPFVRLIRPKKFBGAGKATTFANRILWLDYADAANKGSEVTLMDGNALVEIK 592
Db 544 APQEPKTIEMKPKHKNPANGSEKVIYKYDIAVDKDDADVIVNDEVTLMDGNALYITK-K 602
Qy 593 VESGVITELVGEILBESVTKTKLITWLADIELVPLSIVEFDYILSKKLEBEDFDLND 652
Db 603 NDDG--SMVAKNLWLBGFKTKKHLTWLADTKQVVVPDVLVDFHLITKRLBEDSEFED 659
Qy 653 NLNPTREIRIPALGDANMNIRKGEIIOLEKGYRCDAPFIRSSKPVULFAIPDGR 709
Db 660 FLTPQTEHFDALINDLVKMDKIDIIQFEKGYRRLDA-LPKQKGFVVFETIPIDGK 715

RESULT 3
US-10-032-585-7526
; Sequence 7526, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent version 3.1
; SEQ ID NO 7526
; LENGTH: 725
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7526

Query Match 41.2%; Score 1537; DB 12; Length 725;
Best Local Similarity 47.3%; Pred. No. 1.4e+135; Indels 22; Gaps 9;
Matches 303; Conservative 121; Mismatches 195;

```

```

Qy 27 LTINHSLAGSAPTLQFASGESLHGVN-----PIIYIARGASIASLGNIDIEFG-- 77
Db 21 LTIN-----GKAPVAVAEILAAIRVNLAPNSIAIKLVDDKAPAAKLDATETEVFKI 75
Qy 78 -----HYEWLEVAFTL--SGSEFENACLFPVDFGLASRTFLVGHGL--T 118
Db 76 TSKPAALFDNGDEQVAKVMILAQKELVYKNPAKLSQSLFTLDQNLRTFTILG-GLKYS 134
Qy 119 IADIAVSNLAGIQGRWESLAKSKRYONLVVWENSID-----SEKELNVEVNAFVGKR 173

```

```

Qy 81 EMLEVAFTPLSGSEFENACL--FVDGFLASRTFLVGHGLTIDIAVANSN--AGIQOR 134
Db 86 EMIKFALTUKLYVKNFKELAVLDEKLDGHLNRSRFTVGVYSLADIAIGVLRANALMG-- 143
Qy 135 WESLRSKRYONLVVWENS--DSEYKEALNEVVAAPVGRKGIGKS--PAPSLKEKYHD 189
Db 144 --SVYKNGVYANISRWYLLADDKRFEGSVELMTKSLAEVRAKKAASAKTAAGAAGKEAH- 200
Qy 190 SKDPSAPREVDPGAKVGVYCRFAPBPSGYLHIGAKAALLNKYFAERYOGRILYRFPDT 249

```

```

Db 201 ---KASFEIDLPGAMGKVTTRFPPEPSGYLHIGHAKAVALNEVFAHFKGLIIRFDDT 257
Oy 250 NPSKSNFEVENLKDIEITLGIKYDAVYTSDFPKLMEAMESLIKOKAYIDDPKQOM 309
Db 258 NPTKEKEFODSIIEDELGLIGKQDQITSSDYFOTMTDLAVKMKDKNAACDDTPVDTM 317
Oy 310 RKEKMDIESRCRNNTVEENLSLW-KENVNGTERGMOCVCKGLDMODPNKSLRDPVYR 368
Db 318 REGVAVGASARRESVVENLRIPTFEEMKNGTEBGLKNCILAKIDYKAPNKALRPVYR 377
Oy 369 CNTDPHHRVSKKYKYPFYDPACPFVDALLEGVTHALRSSEYHNRNAQYRILQDMGLRRV 428
Db 378 CNLTPHRTGTETMKYPIYDFCVPVDSIEGVTHALRINEYDRRNPQYEMMQKTLGLRPV 437
Oy 429 EIVFESRLNMYTTLISKRLKLMFVONKVEBMTDRPFTVOGIVRGLKVALLOFLIOQ 488
Db 438 AIMPGRNPFRTLLSKRKLQMPVDKNVSNWDDRFPTVGRGVRRTGELRNFIISQ 497
Oy 489 GASKNLMEWDKMTINKKIIDPVCARHTAVLKQRYIFLTJNGPEEPFVRIIPRHKKF 548
Db 498 GPSNIIINLEWSVIAAMKIIIDPAPRFTAVDAGNVSVKLLNGKPEYTESKPKHKN 557
Oy 549 EGACKKATTFANRMLDYADAAAINKGEVTLMDMGNAIYKEIKVESGITELVGEHLLE 608
Db 558 PEVGNKDVIFADQVLIQED-ADLTGESEVTFMDGNIIVSGKNKEGDIIVKSEANLHLE 616
Oy 609 GSVTKTKITWMLDIEELVPLSLVEFPYLSKKLEDEDEPLDNLNCTREIIPALODA 668
Db 617 GDFERTSKKITWLVYTKDKVEIDMDFDLITKDLNDNNEDEPITETEPHTKGFADL 676
Oy 669 NMRNIKGEIIOLEKGYRCAPFIRSSKPVLPFAIPDGR 709
Db 677 NVGKLKAGDIIQFERKGFYRVDKP-SEGGKPAVLTYTIDGK 716

```

RESULT 4

```

US-10-128-714-3375
; Sequence 3375, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Mengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3375
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3375

```

Query Match 39.4%; Score 1471.5; DB 15; Length 715;
 Best Local Similarity 43.0%; Pred. No. 26-129;
 Matches 310; Conservative 128; Mismatches 234; Indels 49; Gaps 14;

```

Oy 4 ALSEKOSPPISTICAALVGLPLTINSLAAGSAPTLQF--ASGESLHGVNPILLYIAR 61
Db 26 SINARBPSPVISTYEDTAV-----LBEQKAVQYIGASGNPFLGLINAVQELRK 76
Oy 62 GASIASLGGKND-IEFGHVEWLEVAPTF--LSGSEFNACLPVDGFLASRTFLVGHGT 118
Db 77 --DEPFLNSKQEKLE---NEWLSQLEAFAPLDKALDPRIQRLDTHLLSFFVGVALS 130
Oy 119 IADIAVMSNLAGIQRMESLRKSKKYQNLVWFPMSID-----SEYKALNEVVAAPFGK 172
Db 131 TADIALMGARLGNVAAVAIAIKG-SLVVNTWTFYFLBLCWMASTTEVLNQAVR----- 184
Oy 173 RGIGKSPASLKEKYKHDSKDSPEVDLPKAGVGVCRFAPSPSGYLIHGHAKAALNK 232
Db 185 -----EKKAARAKGASVDIALINTEKG-VTRFPPEPSGYLIHGHAKAALND 232
Oy 233 YFA-ERYQRLIVFPDDTNPESKSENFVENLKDIEITGIKYDAVYTSDFPKLMEANAE 291
Db 233 YFAHEKXNGTLVAFDDTNPSEKLEFQDAIIEDLALMGIKPKDKSYTSDFDELQYAL 292
Oy 292 SLIQGKAYIDTPKEQMRKRMGIESRCRNNTVEENLSMKEMVNGTERGMOCVARG 351
Db 293 QIIXDGAAYADDTKEVVAEQRMNGKPSKRDAASVEENLARFBEKKCTPGLRWCTIRAK 352
Oy 352 LDMODPNKSLADPYVYRCNTDPHHRVSGKYVFPYDPACPFVDALLEGVTHALRSSEYHD 411
Db 353 MSVDPNKAAMDPIYIRCNPAHPHRTGTMKIYFYDPACIYVDSIEGVTHALRTIEYRD 412
Oy 412 RNAQYRILQDMGRARVEIYFESRLNMYTTLISKRLKLMFVONKVEBMTDRPFTVOGI 471
Db 413 RNQYQWFLDTLKLKRNHQIWMOPAMNPIRTLLSKRLTKLVNQGVMWDDPRFTIRGI 472
Oy 472 VRGKLYEALIQFLQOGASKNLMEWDKMTINKKIIDPVCARHTAVLKQRYIFLT 531
Db 473 RRRGMTIPALREFLIKQSPSKNITNLDWTLIWTATKKYIDVAPRHETAILKDKWKAIVK 532
Oy 532 NGPEEPFRILIPRHKKEGAKKATTFANRMLDYADAAAINKGEVTLMDMGNAIYKEI 591
Db 533 GGPATPYTEKPKGKAPVAMKRVFGNTVYIPEQDDSKSKODEITLMSMGNAIVAKI 592
Oy 592 KVE--SGVTELVEGLHLEGSVTKTKITWLA-DIEELVPLSLVEFPYLSKKLEBDE 648
Db 593 ETDPTSGIVLELELHLEHGFKTEKKVTLSTEGDILVELVDPYLVNKTLOQDD 652
Oy 649 DFLDNLNCTREIIPALGDANMRNIKGEIIOLEKGYRCAPFIRSSKPVLPFAIPDGR 708
Db 653 VLEDVLMNTEFRBDAAVADCNVALKEGDIIOFERKGYRVDRAYV-PGKPAVLFIPIPTG 711
Oy 709 R 709
Db 712 K 712

```

RESULT 5

```

US-10-128-714-8375
; Sequence 8375, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Mengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066

```

```

: PRIOR FILING DATE: 2001-04-27
: PRIOR APPLICATION NUMBER: US 60/295,890
: PRIOR FILING DATE: 2001-06-05
: PRIOR APPLICATION NUMBER: US 60/303,899
: PRIOR FILING DATE: 2001-07-09
: PRIOR APPLICATION NUMBER: US 60/316,362
: PRIOR FILING DATE: 2001-08-31
: NUMBER OF SEQ ID NOS: 8603
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 8375
:
: LENGTH: 715
:
: TYPE: PRNT
: ORGANISM: Aspergillus fumigatus
US-10-128-714-8375

```

Query Match:	39.2%;	Score 1465.5;	DB 15;	Length 715;
Best Local Similarity	42.9%;	Pred. No. 7.4e-129;		
Matches 309;	Conservative 128;	Mismatches 235;	Indels 49;	Gaps 14;

Oy 4 ALSPKSDPPIISICAAKVGLPTINHSIAAGSAPTLGC--ASGESLHGVPITLYIAR 61
:::
Db 26 SINARPSPVISITYEDTAV-----LRGDAVAVQYTGASNPFIPLINAVOELRK 76

```

62 GASIASLSSKND-IEFGHVEMLEAPTF--LSCGSEFNNACLFVDFGLASRTFLVGHGLT 118
      |      |      |      |      |      |      |      |      |      |
77 --DPEFLNSKDEKLE---NEMLSQLEAFPLDFFKALDPELQRLDTHLLLSRFVVGVALS 130

```

[illegible]

Db -----EKAARAKAGASYDIALNTKEG--VTRPPPESSGYLHIGAAKALLND 232

OY I 13 KGIQSKSPASLKERKHDSKDPSAEVDLDPAKVGVCVRRAPEBGGYLHIGAAKALLNK 232

233 YFAHEKYNQITLLVRFDITPNSNEKLEFODAIIEDLTALMGKIKPKDKSYSTSDYFDELTQYVAL 292

2293 QIIDGNAYADTEKEVMAEQRNGGKPSKRDDASVEENLARFEEMKKKGPEGLRWCIRAK 352

Db

353 MSVNPENKAMRDPVIYRCNPAPHHRTGTYKKIIPYTFDACPVDVSIEGVTHALRTIEYRD 412

```

413 RNFOYQWELDTLKLNRHVQIWDFAKNWFIRLTLSKRLTKLVNGGVWGMDSPREPTIRGI 472

```

```

473 RRRGTTPLALREFLIKQSPSKNITNDWITIMATKKYIDPVARPHALIKKDWKAIYV 532

```

Db 533 GGPATPYITTEKPKIGKKNPAVGMKKYVFGNIVIFPOEDALSFKODEBITLMSGNAILVRKI 592

Db
593 ETDPDTSGLKELBLELHLESGDFKTEKKVYIWLSTFGQDLIPVELVDYDFYLKINKDTLQEDD 652
650 DEFDWINDGMBRFDYDQANMAYRTPGQDTSPQVYVQCDADPTDSCGPDINT EATDNC 708

Db 653 VLEEDLNKATFREDAVADCNVAELKEGDIIFERKGYRVDRAYV-PGRPAVLNIPFG 711

Db 712 K 712

```

US-10-369-493-13278
; Sequence 13278, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13278
; LENGTH: 606
; TYPE: prt
; ORGANISM: Aspergillus nidulans
US-10-369-493-13278

```

Query Match	39.1%;	Score 1459.5;	DB 12;	Length 606;
Best Local Similarity	47.9%;	Pred. No. 2.1e-128;		
Matches 290;	Conservative 106;	Mismatches 195;	Indels 15;	Gaps 7

Qy 106 LASTETVLGAGLTADIAVMSNLAGIGOMESLEKSKYONLVAMFNSLDSEYKALNEV 165
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 6 LIRSFVVGYALSTADIALLMGALRGNRVAIGTIRK-ETLITLPWFNFPLE----ELCPWA 60

DB
QY
166
61

VAAALQALNNAKDD-----KRAAKAKEGASYDIALTLNTDKG-VVTRFPPEPSSGYLHIGHA 113

QY	426	KAALLINKTFA-EXTGKRLVRFEDINFESLANSNEV	ENLNDLLELSJSLIKLIDVLA	115SDIF	208
DB	114	KAALLINDYFAHEKYGKTLVRFDDTNPSNEKOEFO	DALIEDLALMGIKEDKTSYSSDYFO		173

DB 174 ELQOYAVQIIKDGNAYADDTDKETMASQCRNGGEPSCRKRDATVEENLAREFEEMTKTBECA 233

234 RWCRRAKSVNDENKAMKDPVIATRCNPAPAHRTGKTKAYPTIYFCFPIIDSLSECVTHAL 293

294 RTLEYRDRNPQYQFLDITLKLNRVQIWDPARNNFVRTLSKRKLTLYLGGVGVGMSDPR 353

Db

354 PFTIRIGRRGWTIPALREIFILQSPSRNI VNLDMTSPATNKKYIDPAAPHYAVLKCD 413

Db
414 AVAKTIKGGPDSPTQDKRKHNKNPDVGFKVATYSSSILFPDQEDAKSFQKDEETILMNWG 473
608 NATIETETRE -GATITETVCEUUIPCSLTETTKRTMYADIEPLVLSLVEFDDYLISKKK 643

Db
474 NAFVRKIYTDASGVVTELELDLINEKGDFKTEKKVTWLTSTDDQDLPVPVDIVDPFLINKDS 533
:::
:
:
:

644 ISEPPDEPI DNTNCCMPBPTPATGNAAMNNIKRGGTIOLEPKGYVBACDAPETPSSKVATF 703
:::
:
:
:

```

Db      534 M Q E E D S L E D V L N P K T E F F H E D A V A D S N V A L L A E G D I I Q E R K G Y R L D - K V A S P G K P A V E F      592
Ox      704 A I P N G R 709

```

Db 593 NIPSGK 598

RESULT 6

RESULT 7

Query Match	36.1%;	Score 1349;	DB 12;	Length 618;
Best Local Similarity	46.0%;	Pred. No. 5.8e-118;		
Matches 285;	Conservative 100;	Mismatches 210;	Indels 24;	Gaps 10;

Query Match	35.9%	Score 1339;	DB 12;	Length 1149;
Best Local Similarity	39.5%;	Pred. No. 1.4e-116;		
Matches 292;	Conservative 132;	Mismatches 243;	Indels 72;	Gaps 13

```

QY 2 EALISFKDPSPIISOAKLVGLPLTINSLAAGSAPTLOFA-----SGSSLHG 52
Db 6 ELVKANREOPPYASIALAASGFSL-----EKSVQFSKOOLALUDGILLSND 55
QY 53 NPILLYARGASIA-SLSGRNDIEFGHVWVLEAYP-----TFLSGSEFENACLFPVD 103
Db 56 VELARIIAQSDPADASLLGSSIIDFAVVDLDTLNIAVTKNDVSLLLGKQFP----- 108
QY 104 GFLASRFLVGHGLTIDIAVWSMLAGIQORWSLRKSKYQNLVRFPNSIDSEYKEALN 163
Db 109 -----FKIINDSLTVADFAIFIS-----VAHNNPOLK-----AKFSGIIDKVLKEPTL 150
QY 164 EYVAAPVGRKGIGKSP---APSLKEKHDSKDPAPEVDLPGARKVGVCFAPPEPSYL 220
Db 151 AAHNHFGLYKSAAPATYASTGKEK---KQDEGKFVELRGARKGKVVAFPPASGYL 206
QY 221 HIGHAKAALLINKYFAERYOGRLLYRFDDTPNPSKESNEFVENLKDIEYLGICYDAVYTS 280
Db 207 HIGHAKAALLNYYQCAFEGOLIMRFPDTPNPAKENAFENHIKEDLSMLNIVPBRWTHSS 266
QY 281 DYFPKLEMAESLTKOGKAYIIDTPKEOKMRKMDGIESRCRNNTVEENISLYMKENVNGT 340
Db 267 DHEFMILLTMECKILKEGKAFVDDTDTETMERNERQODRSRRSNTPEKNILQIMEEMKGS 326
QY 341 ERMGQCCVRKLMODPNKSLYBDVVYRCMTDPNHRGSKKKYPTDPAFCFPDALGCV 400
Db 327 PKGLTCCVRKIMDKSNNGAMRDPTIYRCRPEEYVRGLKKVPTDPTFCPIVDSEGV 386
QY 401 THALRSSEYNDRNAOYRILODMGELREVEIYFESRLNMVVTLSKRLKLFVONKKVEDM 460
Db 387 THALRTTEYHNRDQYFICDALGRLRPHIWEYARLNMVTYVMSKRLTLWVDEGHVGM 446
QY 461 TDPEEPTVOGIVRGKLEVALIOFILOOGASKNINLMEWDLKWTINKKIIDPVCARHTAV 520
Db 447 DDBRLFTVBRGVMRRGLTVEGKQETVAOGSSRSVMMEMEDKIMAFNNKKVIDPAARYAL 506
QY 521 LKQORVI-FTLUNGREPEFVRLIRHKHKEPEGAGKATTPANRIMLVDADAANAINKGEVY 579
Db 507 DSTSPSLVIELTUSISDTSNV-SLHFKNABIGSKDVHKGKRLLEVDADAALKEGETVY 565
QY 580 LMDMGNAIVEIKESGVITELVGEHLLEGSVKTTKTKITWLMADIE---ELVPLSLVEF 635
Db 566 FVMMGNIKIKGIEKGAIVTIKISATLODMTYDKTKTKVLMGSVKAAGKTIIVVYADY 625
QY 636 DYLSKKKLEDEBDFLINDNCTRRREIYALGDANMRNIKGEIIOLERGKYRCDAPEIR 695

```

Db 626 DHIISKALIGKDEEMKQFINPDSVHYTMMGEPAIKVYKGDIIQIQKRGFYVDQFINP 685
Qy 696 SSK-----PVLFPAIPDG 708
Db 686 KSELGSETPFLIIAIPDG 704

RESULT 9
US-10-104-047-3100
; Sequence 3100, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104.047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3100
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3100

Query Match 34.5%; Score 1287.5; DB 12; Length 717;
Best Local Similarity 41.2%; Pred. No. 4,7e-112;
Matches 280; Conservative 127; Mismatches 244; Indels 29; Gaps 13;

Qy 46 GE-SLHGVNPIILIYARGASIASLGGKNDIEFGHVENLEVAPTFLSGSEFENACLFVDG 104
Db 43 GETIKGEDAICDYFISKCADQSLGCDAB--TVKEWRHIASTTTKENVLEQ-IEKLDI 99
Qy 105 FLAARTPLVGHGLTADIAVMSNLAGIQRMESLRKSKYQNVLRWFMSIOSEYKEALNE 164
Db 100 YIARTYIVRVFTIADISMAAI-----KLTVEIIPAKCTPASPWYIYA--IPSKQ 152
Qy 165 VVAAPVGRGIGKSPAPSLKEKVH-----DSKDPAPAEVDLPGAKVGVCFAPAP 216
Db 153 TTGAVGSLRMAAQAQAPKKEIATKKEISLDGPSEMP--LPAPMGVYTRFPPEA 210
Qy 217 SGYLHIGAKAALLNKYPABRYOGRLIIVFPDTPNSKESNEFENLLDITLGIKYDAV 276
Db 211 SGYMHIGHVAKAMLYYAKRYOGKLIIRFDTPNSKEKEEETSIIIDLAKGIKADLF 270
Qy 277 TYTSDYFPKLMEMESLKGKAYIDTPKQOMRKERMDGIESRCRNTVEENLSLWKEM 336
Db 271 SHTSDYFPVILIDYARQMIREGIAFMDNDQETMRKERMERKESKLRNTSPENLRLEBAL 330
Qy 337 VNGTERGQCCVRGLDMQDPNKSRLDPVYRCNTDPHHRVGSKYKVPTYDPACFPVDA 396
Db 331 CRGEPEVODYCLRAKIDKSDNGTIRDPVLYRFPYLTMRGDKKAYPCIDLACPIYDS 390
Qy 397 LEGVTHALRSSEYHNRNAQYRIIDMGIRVYEIYFESRLMNVYTLSSKRLMFWQKK 456
Db 391 IEGLYTHARTTEYKORDQYMWIIOKALRLRPVHLVEFARLNFQYTLMSKRLTMVVEHNE 450
Qy 457 VEDWTDPRFPYVQGIIVRGKLYEALIOFTLQOGASKNLNLEMDLMTINKKIIDPVCA 516
Db 451 VDGDNDPFPYVQGIIVRGKLYEALIRAFILSGFSEKRVVVTMWDFFWSDNKKILERNALR 510
Qy 517 HTAVLKQDQVIFTLNPGPEEPFRILPRHKKKEGAGKATTPANRIMWDYADAALINKGE 576
Db 511 FMGVAKSPFVTLTLN-VQPGFIK-V-PNHPKDESQVSDIAGPVMLEFRDADQGLKGE 568
Qy 577 EVTLMDGNAIVKEIKVES-GVITELVEELHLEGSVTKTKITWLADIIBELVPLSLVEF 635
Db 569 EFTLMRGSAKYTAIARADAGAVTGTGEFVNDGPKQT-AKVNVVAACPEHNEHTVVEY 627
Qy 636 DYLISKKLLED--EDFLDNLNPTCTREIIPALGDANNRNIKRGIIQLERKGYRCAP 692

Db 628 DDLTKTLEBTDNFEEDFLTSKDHPTQMEYTVIYNACKGLKENQIVQLERGRGYRCDP 687
Qy 693 FIRSKFVLPALPDGROQA 712
Db 688 DTENS-PAKLILIPDGKKA 706

RESULT 10
US-10-369-493-751
; Sequence 751, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369.493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 751
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-369-493-751

Query Match 20.2%; Score 753; DB 12; Length 554;
Best Local Similarity 35.2%; Pred. No. 8,7e-62;
Matches 189; Conservative 91; Mismatches 205; Indels 52; Gaps 14;
Qy 199 DLPAKGVGVVRAPEPSGILHIGAKAALLNKYPABRYOGRLIIVFPDTPNSKESNEF 258
Db 19 DLASGKTTVHTRPPEPENGILHIGHAKSICLNGIADYKQCNLRFDDNPVKEDIEY 78
Qy 259 VENLLKDIEITLGIKYDA-VTYTSDYFPKLMEMESLKGKAYIDTPKQOMRKER--- 313
Db 79 VESTIKNDVENWGFHMSGNVRRSSDYFDQLAHAYIELINKGLAYVDELPEQIRREYRGLT 138
Qy 314 MDGIESCRNNTVEENLSLWKEM-VNGTERGQCCVRGLDMQDPNKSRLDPVYRCNTD 372
Db 139 QPGKNSPYRDRSVEENIALFEKMRAGGFEBS-KACLPRAKIDMASPFIYMRDPVLYRIKFA 197
Qy 373 PNHVGSKYKYPYDYDPAACFPVDALBGVTHALRSSEYHNRNAQYRIIDQWGLR-RVEIY 431
Db 198 EHHQTNKWCICYPMYDFTHCISDLBGLTHSLCTLEFODNRRLYDWDLDNTITIVHPRQY 257
Qy 432 EFSRLNMYVYTLSSKRLMFWQNKVVEDWTDPRFPYVQGIIVRGKLYEALIOFTLQOGAS 491
Db 258 EFSRLNLEYITMSRKUNLNLVTDKGVESGWDPRMPTISGLRRGYTAASIREFCRIGVT 317
Qy 492 KNLNLEMDKIMTKINKIIDPVCAHHTAVLKDQGVITFLNPGPEEPFRILPRHKKKEGA 551
Db 318 KQDWTIEAASIESCIREDLNENAPRAMAVIDPVGLVIENYQGE--NVTMPPNHNKPEM 375
Qy 552 GKATTPANRIMWLYAD-AAALNK-----GEVTLMDGNA-IVKEIKESGVITELV 602
Db 376 GSRVPPSGELWIDRPFREANRQYRLVLGKEVRLR--NAVYIAERVEKA----- 427
Qy 603 GELHLEGSVKT-----TKLK--ITWLADIIBELVPLSLVEFDYLISKKK 643
Db 428 -----EGNITTIIFCTYADLTISKDPAQKAKYGVINWVAHAL-PVEIRIYDLRFSPVN 481
Qy 644 LEEDEDFLDNLNPTCTREIIPALGDANNRNIKRGIIQLERKGYRCAPFIRSKPV 700
Db 482 PGAADDFLAVINPESLVIKQGFAPSPSLKDAVAGRAFOFEREGYCLDSRHSHTAEKPV 538

RESULT 11


```

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13759
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13759

Query Match      18.9%; Score 704.5; DB 12; Length 555;
Best Local Similarity 34.8%; Pred. No. 3.3e-57;
Matches 189; Conservative 87; Mismatches 206; Indels 61; Gaps 18;

QY 197 EVDFGAKGVKVCVFAPEPSGYLHIGAKAALNKYFAERYOGRILVFPDTPNPKESKN 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 17 QADLDGSGHGTQIVRFPPEPNGLIHGAKSICVNFGLAOFEGVTHLRFDDTPNPKEDQ 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 EFVENLKDITLGIKYDA-VTYSDFPKLMEASLTKOGKAYIDDPKEQMRKER- 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 77 EYDIAESDVKMLGFEWSEGEVRYASQYFDQLDMVAVELIKAGKAYVDDLTEBOAKEYRGS 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 --MDGIESRCNNNTVEENLSLMEKENVNGTERGMQCCVKGKLDMDPNKSLRDPVYRCNT 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 137 LTERGKNSPFRDRSVEENLDMFARRAGEFPDGAFLAKIDMASPNMLRDPVYRIRH 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 372 DPHHRVSGKYKVPYTPDFACEFVDALLEGVTHALSSSEYHDINAQYRILIDMGL-RVYEI 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 197 AHHTQGTGKCLIPRYDFTHQSDALBEGITHSICLFEFSHRPLYEMWLDLALPVARRQ 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 YEFRLNNVTLLSKRLKLMFVQNKKEVDWTDPRFTVOGIVRGLKVEALLIQFLQOGA 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 257 YEFRLNNVTLLSKRLKLMFVQNKKEVDWTDPRFTVOGIVRGLKVEALLIQFLQOGA 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 491 SKNLNMEMDKLMTINKKIIDPVCAHRTAVLKDQVIFLTNGPEEPVRI-LPHKKFE 549
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 317 NRSQGVDFGMLFESIRDDLANARAWCVLRPLRV--ITNPPDQVENLELPHPKPE 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 550 GAGKATTPANRIMLDVAD-----AAAIKGEVTLMDGNAIVK--EIKVESGVI 598
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 375 ELGTRQLPFAEIRIYDRDPEEPKGYKLEPNEEVALR--GSVVIADAELIKADANI 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 599 TEL-----VGBLHLEGSVKTKLITWL--ADIEELVPLSLVEFDYLI--SKKKL 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 433 VELRCSYDPTLTK-NPEG--RKVGVIHWPAASICEVAL--YDRLFRSPNPKA 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 645 EDEDFLDNLNLP-----CTREIPALGDANMNIRGELIOLERKGYRCDAPIRS 696
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 486 EDSASFLLDNLNLPGLQLVLTGC--RAEPLGNAQPE-----DRFQEREGYFVAD--TKD 535
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 697 SKP 699
   |||
DB 536 SKP 538

RESULT 14
; Sequence 412, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

```

```

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 412
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-412

Query Match      18.8%; Score 703; DB 12; Length 554;
Best Local Similarity 33.5%; Pred. No. 4.6e-57;
Matches 177; Conservative 94; Mismatches 204; Indels 54; Gaps 13;

QY 199 DLPAKGVKVCVFAPEPSGYLHIGAKAALNKYFAERYOGRILVFPDTPNPKESNEF 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 19 DLGKHTSVTRFPPEPNGLIHGAKSICVNFGLAOFEGVTHLRFDDTPNPKEDIEY 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 259 VENLKDITLGIKYDA-VTYSDFPKLMEASLTKOGKAYIDDPKEQMRKER- 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 79 VESIKQDVQMGFPQMSGDIRSSDYFDQLHQYALIELINKGLAYVDELSPDEIRYGTLK 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 314 MDGIESRCNNNTVEENLSLMEKENVNGTERGMQCCVKGKLDMDPNKSLRDPVYRCNTDP 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 139 EPGKNSPFRDRSVEENLDMFARRAGEFPDGAFLAKIDMASPFMWRDPLVYRIKFAE 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 374 HHRVSGKYKVPYTPDFACEFVDALLEGVTHALSSSEYHDINAQYRILIDMGL-RVYEI 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 199 HHQGNKMKCIPLYDFTHCISDALEGITHSICLFEQDNRRLYDWNLTITDCHPROYE 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 FSRNLNVTLLSKRLKLMFVQNKKEVDWTDPRFTVOGIVRGLKVEALLIQFLQOGASK 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 259 FSRNLNVTLLSKRLKLMFVQNKKEVDWTDPRFTVOGIVRGLKVEALLIQFLQOGASK 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 493 NLNMEMDKLMTINKKIIDPVCAHRTAVLKDQVIFLTNGPEEPVRI-LPHKKFE 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 319 QNNVEMAALSCIRDDLNEHAPRAMAVLDPVKLV--IENPBESEVILTMNHNPKPEMG 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 553 KKATTPANRIMLDVAD-----AAAIKGEVTLMDGNAIVK--EIKVESGVI 603
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 377 SREVPFSEELIDRADPREEANROYKRLWNGKEVRLR--NAVYVKKERVKDA----- 427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 604 ELHLEGSVKLT-----TKLK--ITWLADIEELVPLSLVEFDYLSKKKL 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 428 ----EGNITTYCHYDQGTINKDPADGRKVGVIHWSAV--HAAPAEFYLDRLFSVNP 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 645 EDEDFLDNLNLPCTREIPALG--DANNRIKGEITIIOLERKGYRCDA 691
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 483 GAEDDFLSTLN--TESLVVRKGFVEGLVNMALKEITYQFEREGYFVADS 529
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
; Sequence 9393, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9393
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9393

Query Match      18.3%; Score 684.5; DB 12; Length 573;

```

Best Local Similarity 33.3%; Pred. No. 2.7e-55;
Matches 179; Conservative 86; Mismatches 217; Indels 55; Gaps 10

```

0Y 211 RAPPESSGVLIHGHAAKALLINKKFAERYOSRLVREDPTNPSKESNEFVENLKJOTETLG 270
Db 32 RPPPEBNGVLIHGHAAAIICIDPFVAABFGCHCTLRDNTNPSKEDPAFAAIDEDVSWLG 91
0Y 271 IKYDAVTVSDYFPPKLMENAESLIIKOGKAVIIDTTPKEOMKER-----MODISEGRCNNTV 326
Db 92 FHMNMLRHTSDVFEVUYLAELAKLIDAGKAYVCDLNEQVREHYGTLTEAGRBPFPMBERSP 151
0Y 327 BENLSLWKEWVNGTERGMOCCVARGKLDMODPNKSLRDPVYYRNCNTDPHHRVSGSKXYKYP 386
Db 152 DENLELFROMRAGTFPDGTRTLRAKIDMASGNINLNDPALYRIKHVHQGTMTWPIYPM 211
0Y 387 YDFACFPVDALLEGVTHALBSSEHEDHNAOYRILDOMGLRVEI-----430
Db 212 YDFALSDAIGIHTSLCTLEBED-----HREPLYMCINHVVILPNNSHLLKLEPLDKGFP 266
0Y 431 -----YEFSLMNVTVLLSRKLIWFOUNKKVBDMWDPRPPTOGIYRGLAKYEALI 482
Db 267 QEPSPROTEFSRLNINITYMSRKLTALVDEKLVGCMDDPRMYTIOGLERRRYTPAMR 326
0Y 483 QPILOGASKNLMLMEWMDKLTINKKIIDPVCARHTAVLKDORVIFLTITNGPSEPFVR-1 541
Db 327 LEVERHIGISKONSIIDPSVLNENCRENDLTIAARRAATIAPMKV--LTLPEHDEQOLI 384
0Y 542 LPRHKKFPBAGAKKATTPANRIMLDVADAAIN-----KGEVTLMDGNAIKVE-1K 592
Db 385 FPNHKKDPTOGTRTVPFSRELTIERDDFSEVPKGMKRLVPGGEVRLRGAGIARIDEVVK 444
0Y 593 VESGVITELVGLH-----LEGSVYTKYTKLTWLDIEBLVPLSLVERPYLISKKULEE 646
Db 445 NAEGHIALHGLWLDTPSRDGMEGANHKVGGTJHWS-APHAVAARILYDLRFSIEKPPD 503
0Y 647 DED-----FLDNINPCTRREIIPALGDAMRNKIGGEIIQLERKGYRCDAPIRBSKXY 700
Db 504 NTDGKTYRDFLNPDSKRVVHGVIETPPAAQOTABENHAFQEBELGIVFTDRHDHDHDTHEV 560

```

Search completed: January 25, 2004, 17:38:18
Job time : 248 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 16:58:19 / Search time 71 Seconds
(without alignments)
968.459 Million cell updates/sec

Title: US-09-831-683B-10

Perfect score: 3734

Sequence: 1 MEALSFKSDSPISIIICAA.....SSKPVLFALPDGRQASLS 715

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: PIR 76:*

1: p1r1:.*
2: p1r2:.*
3: p1r3:.*
4: p1r4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2657	71.2	719	2 T52043	probable glutamate
2	2644.5	70.8	728	2 T01200	probable glutamate
3	1574.5	42.2	716	2 T37830	probable glutamate
4	1562	41.8	724	1 S53934	glutamate-cRNA lig
5	1498	40.1	1714	1 S18644	multifunctional am
6	1437	38.5	1440	1 SYHUOT	multifunctional am
7	1339	35.9	1149	2 T27567	hypothetical prote
8	757.5	20.3	555	2 AC0585	glutaminyl-cRNA sy
9	753	20.2	554	1 SYEQOT	glutamine-cRNA lig
10	747	20.0	554	1 F90717	glutamine-cRNA syn
11	747	20.0	554	2 F85567	glutamine-cRNA syn
12	743.5	19.9	555	2 AB0321	glutamine-cRNA lig
13	721	19.3	556	2 H82254	glutaminyl-cRNA sy
14	710.5	19.0	557	2 G64118	glutaminyl-cRNA lig
15	704.5	18.9	556	2 P83421	glutaminyl-cRNA sy
16	690	18.5	571	2 A84978	glutamine-cRNA lig
17	683.5	18.3	580	2 G82693	glutaminyl-cRNA lig
18	666.5	17.8	562	2 B81069	glutaminyl-cRNA sy
19	664.5	17.8	573	2 D81799	glutamine-cRNA lig
20	653	17.5	571	2 G75165	glutaminyl-cRNA sy
21	636.5	17.0	794	2 T09643	glutamine-cRNA lig
22	624	16.7	570	2 F71049	glutamine-cRNA lig
23	594	15.9	809	1 SYBYOT	glutamate-cRNA lig
24	589.5	15.8	553	1 H64471	glutamate-cRNA lig
25	575	15.4	852	2 A75253	glutaminyl-cRNA sy
26	573	15.3	786	2 T26811	hypothetical prote
27	572	15.3	811	2 T40275	probable glutaminyl
28	561.5	15.0	775	1 I37422	glutamine-cRNA lig
29	559	15.0	551	2 D69282	glutaminyl-cRNA sy

30	537.5	14.4	553	2 B69167	glutamate-cRNA lig
31	530	14.2	552	2 S65787	glutamate-cRNA lig
32	527.5	14.1	729	2 D86383	probable glutaminyl
33	504	13.5	575	2 S75395	probable glutamine
34	477.5	12.8	544	2 A72459	probable glutaminyl
35	474.5	12.7	586	2 PC4398	glutamine-cRNA lig
36	381.5	10.2	586	2 B84271	glutaminyl-cRNA sy
37	275.5	7.4	485	2 C97022	glutaminyl-cRNA sy
38	271	7.3	473	2 D70405	glutamate-cRNA lig
39	266.5	7.1	468	2 S21172	glutamate-cRNA lig
40	255.5	6.8	464	2 F82019	glutamate-cRNA lig
41	255	6.8	489	2 G72264	glutamate-cRNA lig
42	255	6.8	484	2 S73490	glutamate-cRNA lig
43	250	6.7	463	2 A81337	glutamate-cRNA lig
44	247	6.6	487	2 F72200	glutamate-cRNA lig
45	245.5	6.6	467	2 G84937	glutamate-cRNA lig

ALIGNMENTS

RESULT 1

T52043 Probable glutamate-cRNA ligase (EC 6.1.1.17) [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Jun-2002

C:Accession: T52043

R:Day, I.S.; Golovkin, M.; Reddy, A.S.

Biochim. Biophys. Acta 1399, 219-24, 1998

A:Title: Cloning of the cDNA for glutaminyl-cRNA synthetase from Arabidopsis thaliana.

A:Reference number: 224836; MUID:9765600; PMID:9765600

A:Accession: T52043

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-719 <DAY>

A:Cross-references: EMBL:AF067773; PIDN:AAC36469.1

C:Superfamily: Yeast glutamate-cRNA ligase; glutamine-cRNA ligase homology

C:Keywords: ligase

Query Match	71.2%	Score 2657	DB 2	Length 719
Best Local Similarity	69.0%	Pred. No. 2.4e-169		
Matches	491	Conservative 106	Mismatches 107	Indels 8
Gaps	5			
Qy	5	LSFESKSDSPISIIICAAATVGLPLTINHSIAAGSAPTQFASGESLHGVPNIIITVARGA- 63		
Db	6	LSFEPSPPLSVIALSLASPVITDSSAAATVPSFVSDGRKLGATVLRVGRSAK 65		
Qy	64	SIASLSGKNDIEFGHVVEWLEVAFTPLSGSEFENACLFVDGFLASRTFLVGHGLTIADIA 123		
Db	66	KLPDPFYGNNAFPDSSQIDEDWDVYASFSSGSEFENACGRVDKYLESSFTLVGHSIADVA 125		
Qy	124	VMSNLAGIGQWESLRSKYQNLVRMPNSIDSYKALNEVAAFYGRKIGKS-PAPS 182		
Db	126	IWSALAGTQWRBESLRKSKYQSLVRWNSILDEYSVLNKLVTYV-KKSGGRVAAVP 184		
Qy	183	LKEKVH---DSKQPSAPVDLPGAKYGCVRAPAPSGVHTGHKAALNLYPAERY 238		
Db	185	SKSQQAQVKGQDQKGRPEVDLPEALIGKYLRAPEPSGLHTGHKAALNLYFAERY 244		
Qy	239	QGRLIIVFDDTNPKESNEFEENLKDIEITGICYDAVVTYSDYFPKLMENAELIRQK 298		
Db	245	QGEYIVAFDDTNPKESNEFEENLVKDITGITYEKVTTSDYFPKLMENAELIRMEBK 304		
Qy	299	AYIDDTPEQWRKERMKGIESRCNNVTVEENLSIMKEMVNGTERGMQCCVRGKLDMDPN 358		
Db	305	AYVDDTPEQWRKERMKGIESRCNNVTVEENLSIMKEMVNGTERGMQCCVRGKKNMDPN 364		
Qy	359	KSLPDPYVRCNDTPHHRVGSKYVVPYDPACPFVVALSGVTALSSSEVHDNAOYR 418		
Db	365	KAMDPPYVRCNPMSHRIDGKYIPTYDPACPFVSLBSITHALSSSEVHDNAOYR 424		
Qy	419	ILDDMGLRVEIYFSSLNMYTTLSSRKLLMFQONKKVEDMDTPRPPTVQGIVRGLKV 478		

Db 425 VLEDMGLRQVQLYFERSRLNLTLSKRLMLFVQTGLVDGMDPRFPVTGQIVRGLKI 484

Qy 479 EALIQFIQOGASKNLNMEDKLTNTINKIIPVCAHRTAVLKQDQVIFTLTNGPEEPF 538

Db 485 EALIQFIQOGASKNLNMEDKLTMSINKRIIDPVCPRHTAVARRVLTFLTDGDPBPF 544

Qy 539 VRIIPRHKKEFGAGKATTPFANRIWLDVADAANKGSEVTLMDGNMIVKEI-KVESGV 597

Db 545 VRMI PKHKKFEGAGKATTPFKTSIWLBEADASISVGEVTLMDGNMIVKEITKDEGR 604

Qy 598 ITELVGELHSGSVYTKTKITWMLADIEELVPLSLVEPDYILSKKLEDEDFLDNLNCP 657

Db 605 VTLASGVNLQGSVYTKTKITWLPDTMELVNLITLFDYILITTKKLEDDDEVDVAFVNP 664

Qy 658 TRREIPALGDANMRNIRKGEIITLERKGYRCADAFIRSSKPVLFALPDG 709

Db 665 TKKETLALGDSNMRLKCGDVIQLERKGYFRCDVFPVKSRIYVLSIPDG 716

RESULT 2

T01200
Probable glutamate-cRNA ligase (EC 6.1.1.17) F21E10.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 03-Jun-2002
C:Accession: T01200
R:Davidson, S.; Rohlfing, T.; David, M.; O'Brian, D.
Submitted to the EMBL Data Library, April 1998
A:Description: The sequence of A. thaliana F21E10.
A:Reference number: Z14258
A:Accession: T01200
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-728 <DAV>
A:Cross-references: EMBL:AF058914; NID:G3047074; PID:G3047084; GSPDB:GN00063; ATSP:F21E1
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: ATSP:F21E10.12
A:Map position: 5
A:Introns: 47/2; 89/3; 141/1; 503/3; 659/3
C:Superfamily: Yeast glutamate-cRNA ligase; glutamine-cRNA ligase homology
C:Keywords: aminoacyl-cRNA synthetase; ligase; protein biosynthesis
F:223-499/Domain: glutamine-cRNA ligase homology <EBL>

Query Match 70.8%; Score 2644.5; DB 2; Length 728;
Best Local Similarity 68.2%; Pred. No. 1.7e-168;
Matches 492; Conservative 105; Mismatches 107; Indels 17; Gaps 6;

Qy 5 LSPKSDSPRITICAKLVGLPLTNHSLAAGSAPTLQFAGSESLHGVNPIILYARGA- 63

Db 6 LSPSPSPPLSVTLVLSASPVTTIDSSAAATVPSFVSDGRKLNGLATVLRVGRSAK 65

Qy 64 SIASLSGKNDIEFGHV-----EWLEVAPTFLSGSEFENACLFDVDFLASRTFLVG 114

Db 66 KLDPDYGNNAAPDSQVSLICIMKIDEMVDVAVSVSSGSEFENACGRVDKYLESTFLVG 125

Qy 115 HGLTIADIAVMSNLGICQWRWESLRKSKYQNLVWPNISIDSEYKALNEVVAALVGRG 174

Db 126 HSLSTADVAIWSALAGTQWRWESLRKSKYQSLVWPNISIDSEYKALNEVVAALVGRG 184

Qy 175 IGKS-PASLSKEKHA-----DSKDPAPPEVDLPGAVGVKCVFAPSPSGYLHIGAKAL 229

Db 185 SGKPVAAKSKDSQAQVAKGDQDKGKPEVDLPKAFBEPSSGYLHIGAKAL 244

Qy 230 LNKYFAERYQGLIYRFDITNPKSENEVENILKDIETLGIKYDAVYTSDFPKLMM 289

Db 245 LNKYFAERYQGEVYRFDITNPKSENEVENILKDIETLGIKYDAVYTSDFPKLMM 304

Qy 290 AESLIKQKAYIDTTPKEQMRKEMDGLIESCRNNTVEENISLWKMVNGTERGQCCVR 349

Db 305 AEKLMRBEKAYVDDTPREQMKERMDGIDSKCRNHSVEENILKLMKEMLAGSERGLQCCVR 364

Qy 350 GKLMDQNRKSLARDVYVRCNTDPHHRVGSKYKVPYTDFACPFDALLEGVTHALRSSEY 409

Db 365 GKEMQDPNKAMRDPVYVRCNPMSHHRIQDKYKIYPTDYDFACPFVDSLEGITHALRSSEY 424

Qy 410 HDNRQAQYRIILODMGLRVEIYERSRLNMYTTLSSKRLMLFVQNKYVEDMTDRFPVQ 469

Db 425 HDNRQAQYFVKLEDMGLRQVQLYFERSRLNLTLSKRLMLFVQTGLVDGMDPRFPVTQ 484

Qy 470 GIVRGLKVEALIQFIQOGASKNLNMEDKLTNTINKIIPVCAHRTAVLKQDQVIFT 529

Db 485 GIVRGLKVEALIQFIQOGASKNLNMEDKLTMSINKRIIDPVCPRHTAVARRVLTFL 544

Qy 530 LTNPEEPFVRIIPRHKKEFGAGKATTPFANRIWLDVADAANKGSEVTLMDGNMIVK 589

Db 545 LTDPDEPFVRIIPRHKKEFGAGKATTPFKTSIWLBEADASISVGEVTLMDGNMIVK 604

Qy 590 EI-KVESGVITELVGEHLBSGVYTKTKITWMLADIEELVPLSLVEPDYILSKKLEDE 648

Db 605 EITKDEEGRVYALSGVNLQGSVYTKTKITWLPDTMELVNLITLFDYILITTKKLEDD 664

Qy 649 DFLDNLPCTREIPALGDANMRNIRKGEIITLERKGYRCADAFIRSSKPVLFALPDG 708

Db 665 EVADPVNPTKKTETLALGDSNMRLKCGDVIQLERKGYFRCDVFPVKSRIYVLSIPDG 724

Qy 709 R 709

Db 725 R 725

RESULT 3

T37830
Probable glutamate-cRNA ligase (EC 6.1.1.17) - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Jun-2002
C:Accession: T37830
R:Devlin, K.; Churcher, C.M.; Bartell, B.G.; Rajandream, M.A.; Wood, V.
Submitted to the EMBL Data Library, August 1997
A:Reference number: Z21737
A:Accession: T37830
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-716 <DEY>
A:Cross-references: EMBL:Z98849; PID:CA11515.1; GSPDB:GN00066; SPDB:SPAC17A5.15C
A:Experimental source: strain 972h-; cosmid c17A5
C:Genetics:
A:Gene: SPDB:SPAC17A5.15C
A:Map position: 1
C:Superfamily: Yeast glutamate-cRNA ligase; glutamine-cRNA ligase homology
C:Keywords: ligase

Query Match 42.2%; Score 1574.5; DB 2; Length 716;
Best Local Similarity 49.5%; Pred. No. 4.3e-97;
Matches 319; Conservative 102; Mismatches 201; Indels 23; Gaps 10;

Qy 70 GKNDIEFGHVEMLEVAPTFLSGS-----EFENACLFDVDFLASRTFLVGHGLTIADIAV 125

Db 81 GENDRSL--VESWYETA-SALAGNHNFLSLAQLDDHLIMSLFVGYSLTSDPSIW 137

Qy 126 SNLAGICQWRWESLRKSKYQNLVWPNISIDSEYKALNEVVAALVGRGIGKSPAPSLKE 185

Db 138 GALKSNMMAAGAVR-TCQYFNILARWYKFMDSQ--NAVSVTMEEFKAVNIRK-----KQ 188

Qy 186 KVNDSKDSAPPEVDLPKAGVGVKCVFAPSPSGYLHIGAKALLNKVFARYQGLIYR 245

Db 189 K---SSGENY-EIGLPDIDKQVTRPPEPSGYLHIGAKALNQFFAKYHGLKIVR 244

Qy 246 FDDTNPSEKSENEFENILKDIETLGIKYDAVYTSDFPKLMMASLSIKQKAYIDTTP 305

Db 245 FDDTNPSEKSENEFQDALIEDVALLGIKEDVYVYTSDFITDITHOVCVDMIKSGQAVADDT 304

Qy 306 KEQWRKEMDGLIESCRNNTVEENISLWKMVNGTERGQCCVRGKLDMDQPNKSLRDPV 365

Db 305 VETMRHERTBGIPSKHRRPIEESLEILSEMDKSDVGLKNCIDAKISYENPNKAMRDPV 364

Qy 366 YTRCNTDHNHVGSKYKVPYTDFACPFDALLEGVTHALRSSEYHDNRQAQYRIILODMGL 425

C:Accession: S18644
 R:Cerini, C.; Kerjan, P.; Aebler, M.; Gratecos, D.; Miranda, M.; Semeriva, M.
 EMBL J. 10, 4267-4277, 1991
 A:Title: A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA
 A:Reference number: S18644; MUID:92097547; PMID:1756734
 A:Accession: S18644
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1714 <CER>
 A:Cross-references: GB:M74104; NID:G157563; PIDN:AAA28594.1; PID:G157564
 C:Genetics:
 A:Gene: FlyBase:Acet-qlupro
 A:Cross-references: FlyBase:FBgn0005674
 C:Superfamily: Drosophila multifunctional amino acid-tRNA ligase; amino acid-tRNA ligase
 C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; multifunctional enzyme; protein bios
 F:204-480/Domain: glutamine-tRNA ligase homology <EGL>
 F:755-800/Domain: amino acid-tRNA ligase repeat homology <ATL1>
 F:827-872/Domain: amino acid-tRNA ligase repeat homology <ATL2>
 F:901-946/Domain: amino acid-tRNA ligase repeat homology <ATL3>
 F:980-1025/Domain: amino acid-tRNA ligase repeat homology <ATL4>
 F:1055-1100/Domain: amino acid-tRNA ligase repeat homology <ATL5>
 F:1129-1173/Domain: amino acid-tRNA ligase repeat homology <ATL6>

Query Match 40.1%; Score 1498; DB 1; Length 1714;
 Best local similarity 44.2%; Pred. No. 2.1e-91;
 Matches 321; Conservative 126; Mismatches 247; Indels 32; Gaps 13;

1 MEAALSFSDSPSPISIIICAKLVG--LPITINHSIAASAPITOPASGSL--HGVPRI 56
 1 MSIKKALNNPISGLTAHLINTIVEVEIWSKEET--LQPPDRRLVCHSNNVL 57
 57 LYIARGASIASLGNIDIEFGHVEWLEAPFLSSEFENACLFVDFGLASRTFLVGHG 116
 58 PALAAADYLYGRTAERTQIDHMLSFSLT--CEDDISWALSPLTSPPLPPTVLYANK 115
 117 LTIDIANWSNLAGICGEMESLRKSKYONLVWMNSIDSEKELNVAFAVGKRGIG 176
 116 LTIDFALFENM--HSRYEFALAKGIDPOHVQWYDLITTAQ--PLIQVLOSLEPDAYK 170
 177 KSPAPSLKE--KVHDSKOPSAPEVDLPQAKVGCVAFAPEPSGYLIGHAKALLNKY 233
 171 RSPQSSKEGTAKTGERGQ--EKGFPDLPQAGEMGKVAFFPPASGYLIGHAKALLNKY 229
 234 FAERYGSLIVRPDDTNSKESNEFENLKDIEFLGKYDAVYTSDFPKLMEASL 293
 230 YALVQGLTLMFPDTPAKETVEENYVIGDLQIKPDVFTHSNVFDMIDYCVRL 289
 294 IKCGAVYDTPPKEMKERNKDIGSRNNVNEISLMEKMNNGCTRGMOCCVRGLD 353
 290 IKESKAYVDDTPPEKMKLERQVRVSNRNSVEKNLSLMEEMVGSSEKQNTACAKID 349
 354 MODPKSLRDPVYVRCNTDPHHRVGSKYKVPYTFACPFVDALRGVTHALRSSEYHRN 413
 350 MSSPNCGRDPTIYCKNEPHERITKTKVPTTYFACPIYDAINVTHTLTITTHRD 409
 414 AQYRIIDMGRLRVEIYEFSLNNVYTLSSKRLMVFVONKVEDWTDPRFPVQIVR 473
 410 DQFYFIDALKLRKPYISYSRLNNTNTVLSKRLTWFDSDGLVGMDDPRFPVRIIR 469
 474 RGLKREALIOTILOGASKNLMEWDKMTNKKITIPVCARHNAVLDQGVFTTLNG 533
 470 RGMTVEGKEFLIAGSSKSVFNNWDMKIWAENKVIDIARVYALKEKKVINAVGA 529
 534 PEEFVRI--LPRHKKFEGAGKATTFANRIMLDVDAAINGEVTLMDGNALIVKEI 591
 530 KYE--RIQVSHRPDESLGKKTUULGRRIYIDYDAELKGEVATPLMNCILIKYN 586
 592 KYESGITELVGLHLEGSVTKTKITWLA---DIEELVPLSLVEFYLLSKKLEBDE 648
 587 KDAAGNITSVDALLENKDKFKTKLTLWLAVEDPSAVPFPFCYFNITISKAVLGDE 646
 649 DFLDNLNCTREIRFALGDAMNRNKRGEIITLEKGYRCAPTRRS-----KPYVL 702

Db 647 DFKOFIHKTRDEVPMGLDEPLKCKKGDIIQLQRGFFKVDVAVLPSPGTVNPSPIVL 706
 Oy 703 FAIPDG 708
 Db 707 FSIPDG 712

RESULT 6
 SYNDUT
 multifunctional aminoacyl-tRNA synthetase - human
 N:Alternate names: glutroyl-tRNA synthetase; glutaminyl-tRNA synthetase
 N:Contains: glutamate-tRNA ligase (EC 6.1.1.17); proline-tRNA ligase (EC 6.1.1.15)
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jul-2002
 C:Accession: A38663; S03424; S00969; S38809
 R:Felt, R.; Knippers, R.
 J. Biol. Chem. 266, 1448-1455, 1991
 A:Title: The primary structure of human glutaminyl-tRNA synthetase. A highly conserved c
 A:Reference number: A38663; MUID:91107633; PMID:1980429
 A:Accession: A38663
 A:Molecule type: mRNA
 A:Residues: 1-1440 <EET>
 A:Cross-references: GB:X54326; NID:G31957; PIDN:CAA38224.1; PID:G31958
 A>Note: the cited Genbank accession number, X54327, is not in release 101.0
 R:Knippers, R.
 submitted to the EMBL Data Library, April 1988
 A:Reference number: S03424
 A:Accession: S03424
 A:Molecule type: mRNA
 A:Residues: 96-382, 'I', '384-497', 'IGATSTLQ', '506', 'YT', '509', 'WOME', '514', 'SYL', '518', 'MOSLIWKT', '5.
 A:Cross-references: EMBL:X07466; NID:931769; PIDN:CAA30354.1; PID:9825664
 R:Thoenes, P.; Felt, R.; Schray, B.; Kunze, N.; Knippers, R.
 Nucleic Acids Res. 16, 5391-5406, 1988
 A:Title: The core region of human glutaminyl-tRNA synthetase homologues with the Escheri.
 A:Reference number: S00969; MUID:88262551; PMID:3290852
 A:Accession: S00969
 A:Status: nucleic acid sequence not shown.
 A:Molecule type: mRNA
 A:Residues: 96-146, 'I', '148-191', 'T', '193-382', 'L', '384-416', 'G', '418-497', 'IGATSTLQ', '506', 'YT', '5.
 HO>
 A:Cross-references: EMBL:X07466
 R:Cerini, C.; Kerjan, P.; Aebler, M.; Gratecos, D.; Miranda, M.; Semeriva, M.
 EMBL J. 10, 4267-4277, 1991
 A:Title: A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA
 A:Reference number: S18644; MUID:92097547; PMID:1756734
 A:Contents: annotation; demonstration of glutamyl- and prolyl- tRNA synthetase activities
 R:Kaiser, B.; Eberhard, D.; Knippers, R.
 J. Mol. Biol. 34, 45-53, 1992
 A:Title: Exons encoding the highly conserved part of human glutaminyl-tRNA synthetase.
 A:Reference number: S38809; MUID:92211721; PMID:1556743
 A:Accession: S38809
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 96-382, 'L', '384-463' <KAT>
 C:Genetics:
 A:Gene: GDB:EPBS; QPRS; QARS
 A:Cross-references: GDB:126609; OMIM:138295
 A:Map position: 1432-1442
 C:Superfamily: human multifunctional amino acid-tRNA ligase; amino acid-tRNA ligase repe
 C:Keywords: aminoacyl-tRNA synthetase; ATP; duplication; ligase; protein biosynthesis
 F:125-485/Domain: tRNA-charging <GLN>
 F:126-402/Domain: glutamine-tRNA ligase homology <EGL>
 F:677-733/Region: 57-residue repeat
 F:688-733/Domain: amino acid-tRNA ligase repeat homology <ATL1>
 F:750-806/Region: 57-residue repeat
 F:761-806/Domain: amino acid-tRNA ligase repeat homology <ATL2>
 F:828-884/Region: 57-residue repeat
 F:839-884/Domain: amino acid-tRNA ligase repeat homology <ATL3>

Query Match 38.5%; Score 1437; DB 1; Length 1440;
 Best local similarity 43.5%; Pred. No. 1.9e-87;
 Matches 284; Conservative 133; Mismatches 198; Indels 38; Gaps 8;


```

Qy 74 IEFHGVWLEVAFTPLSGSF-FENACLFVDGFLASRTFLVGHGLTIADIAMVSNLAGIG 132
Db 1 MEHEIDHLEWFSATKLSGSCDSTSTINEINHCISLRTLVGNISLADLCVWATLTKGNA 60
Qy 133 QRMESLRKSKYQNLVRWFNSIDSEYKALNEVVAAPGKKGIGSPASLKEKHSD 192
Db 61 AMOQOLKOKKAPVHVKWPFLEAQ-QAFQSV-----GTRKMDSTTKA 102
Qy 193 PSAPSE-----VDLPAGVGVCVFPAPEPSGYLHIGHAKAALLNKYFAERYOGRLLV 244
Db 103 RVAPEKQDVGKFWELPBAEMKQTVRRPPEASGLIHGHAKAALLNQHIVQNFKGLIM 162
Qy 245 RPDPTNPKESNEFVENLLKDIETLGIKYDAVYTSDFPKLMEVASLTKQKAYIDT 304
Db 163 RPDPTNPKESNEFVENLLKDIETLGIKYDAVYTSDFPKLMEVASLTKQKAYIDT 222
Qy 305 PKQQRKRMGIESRCRNATVEENLSLWKEMVNGTSGMOCVARGKLDMODPNKSLRP 364
Db 223 PAQWKKAREQRIBSKHKNPIELKNLQWMEKKKSGSGHSCCLRAKIDMSNNGCMRP 282
Qy 365 VYRCNTDPHHRVGSKYKVPYDAPCPVDALGVTALRSSEVHDNAOYRRLDMG 424
Db 283 TLVCKIOPHRTGKKNVPTYPDACPVDISBVTALRTTEHDEQYWMITELAG 342
Qy 425 LRARVEIYFSSRLNMYTLLSKRKLMPYQNKVBDWTDPRPPTVQGIYARGLYEALIOF 484
Db 343 IRKPYIWEYSRLNNTVLLSKRKLMPYQNKVBDWTDPRPPTVQGIYARGLYEALIOF 402
Qy 485 ILQOGASKNLMEWDLKMTINKKIDVYCAHRTAVLKQORVIFLTVNGPEPVRILIPR 544
Db 403 IAAQSSSRVNMWMDKIMAFNKKVIDVAPRYVALLKEEVIYVAVPEAOES--MKEVAK 460
Qy 545 HKKEGAGKKAATTPANRIMLDYADAANKGEVTLMGNAIVEI-KVSSGVTELVG 603
Db 461 HPKAPVGLKAVWVSPKVFIEGADSETSEBEMVTFIMWGNLNTKIKHNAIDGKIIISDA 520
Qy 604 ELHLEGVYTKTKITWTLADIEBELVPLSL--VEPDYLSKKKLEDEDFLDNLNPTCRR 661
Db 521 KFNLENDKYKTKTVTWALETTHALPIVICYTHLITKPVLGDEDFEFOYVANKNSGHE 580
Qy 662 IPALGDANMNNIKGEIILQERKGYRCD-----APFIRSSKPVLPFAIPDG 708
Db 581 ELMIGDPLCKDLKGGDIIQLORGRFPCIDQYEPVSPYSCKEAPCVLLIYIPDG 633

RESULT 7
T27567
hypothetical protein ZC434.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27567
R:Wilkinson, J.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z20388
A:Accession: T27567
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1149 <WTL>
A:Cross-references: EMBL:Z27514; PIDN:CA00060.1; GSPDB:GN00019; CESP:ZC434.5
A:Experimental source: clone ZC434
C:Genetics:
A:Gene: CESP:ZC434.5
A:Map position: 1
A:Introns: 47/2; 502/2; 533/3; 649/3; 1110/3

Query Match 35.9%; Score 1339; DB 2; Length 1149;
Best Local Similarity 39.5%; Pred. No. 4.7e-81;
Matches 292; Conservative 132; Mismatches 243; Indels 72; Gaps 13;

Qy 2 EAALSPKSDSPISITICAAKLVGLPLTINSLAGSAPTLQFA-----SGESLHGV 52
Db 6 ELVIVKANREQPPVYSIILALASGFSL-----EKSVQFSEKQALMLNDELISND 55

```

```

Qy 53 NPILLYIARGASIA-SLSGKNDIEFGHVEWLEBAP-----TFLSGSEFENACLFVD 103
Db 56 VEARIITADSTADADSLGSSIIDPAVVDLTNPLAIVTKNDVSLLGKDFP----- 108
Qy 104 GFLASRTFLVGHGLTIADIAMVSNLAGIGQRMESLRKSKYQNLVRWFNSIDSEYKALN 163
Db 109 -----TKILNLSLTVDFAIFS-----VAHNNPOLK-----AKSGIIDKVLKEPTL 150
Qy 164 EVNAAPVGRKIGKSP---APSLKEKHSDSPAPEDLPAGVGVCVFPAPEPSGYL 220
Db 151 AAANPFGVLYSAAPATATSTGK-----KDEGKFWELPBAEMKQTVRRPPEASGYL 206
Qy 221 HIGHAKAALLNKYFAERYOGRLLVPRDPTNPKESNEFVENLLKDIETLGIKYDAVYTS 280
Db 207 HIGHAKAALLNQHIVQNFKGLIMRFPDPTNPKESNEFVENLLKDIETLGIKYDAVYTS 266
Qy 281 DYPFKLMEVASLTKQKAYIDTDPKQQRKRMGIESRCRNATVEENLSLWKEMVNGT 340
Db 267 DHFEMLLTMCHEKLLKEGAFVDDTDTETMRNERQRODSRNSRNTPEKNILQWMEKKGS 326
Qy 341 ERGQCCVARGKLDMDPNKSLRDPVYRCNTDPHHRVGSKYKVPYDAPCPVDALGCV 400
Db 327 PKGLTCCVRMKIDMKNNGAARDPTTYRCPEEVRGLKVKYPTDFTCPIDVSVEGV 386
Qy 401 THALRSSEYHDNRNAQYRIILQDMGLARVEIYFSSRLNMYTLLSKRKLMPYQNKVBDW 460
Db 387 THARTTEYHDNRDQYIFFCIDALGRPHIMEYARLNTVNSKRLTWTVDGSHBGW 446
Qy 461 TDPEPFTVQGIYARGLYEALIOFILQOGASKNLMEWDLKMTINKKIDVYCAHRTAV 520
Db 447 DDPELPYVARGVWRGLTVEGKQFIVAQSGSRVNMWMDKIMAFNKKVIDVAPRYVAL 506
Qy 521 LKQORVIFLTVNGPEPVRILIPRHKKEGAGKKAATTPANRIMLDYADAANKGEVTL 579
Db 507 DSTPLVSIELTDSISDTSNV-SLHPRKNAEISGKDVHKKKLLLEQVDAALKEGELVT 565
Qy 580 LMDGNNAIVEIKIESGVTELVGELHLEGVYTKTKITWTLADIE---ELVPLSLVEF 635
Db 566 FVMGNKIKIGIEKKGAVITKISLTLDNDTDYKTKTVTLGIVKAKAGTIIYVATDY 625
Qy 636 DYLLSKKLEDEDFLDNLNPTCRRRIIPALGDANMNNIKGEIILQERKGYRCDAPFIR 695
Db 626 DHISKALIGDEWDMQFINFDSVHYTKMGEPALIKVYKGGDIIQLORKGYIYDQFPNP 685
Qy 696 SSK-----PVLPFAIPDG 708
Db 686 KSELGVTPLLLAIPDG 704

RESULT 8
AC0585
glutaminyl-tRNA synthetase [imported] - Salmonella enterica subsp. enterica serovar Typh
C:Species: Salmonella enterica subsp. enterica serovar Typh
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0585
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov.
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AC0585
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-555 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05149.1; PID:gl6501922; GSPDB:GN00176
C:Genetics:
A:Gene: STY0724
C:Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology

Query Match 20.3%; Score 757.5; DB 2; Length 555;

```

Best Local Similarity 35.1%; Pred. No. 1e-42;
Matches 189; Conservative 95; Mismatches 201; Indels 53; Gaps 15;

Qy 199 DLPAKVGKVCVRPAPEPSGLYHGHAKALINKYFAERYOGRILVFPDDTPNPKESNEF 258
D 19 DLAGSKHTVTRPPPEPNGYLHGHAKSICLNFGIADYQGCNLRFPDDTPNPKEDIEY 78
Qy 259 VENLKDIETLGIKYDA-VYTSDFPPLKMEMASBLIKOGAKYIDDTPEQMRKER---- 313
D 79 VDSIKNDVEMLGFHWSGDIRYSSDFDQHAHYAELINKGLAYDELTPEDIREYRGLT 138
Qy 314 MDGIESRCRNNTVEENLSLWKEM-VNGTERGQCCVCRKGLDMODPNKSLRDPVYRCNTD 372
D 139 APGKNSPRDRSVEENLALFEKRAAGFPBEG-KACLRKAKIDMASPFYMRDPLVLRIFA 197
Qy 373 PPHRVGSKYKVPYTFDACPVDALGVTHALRSSEYHNDRAOYRRLIDOMGLR-RVEIY 431
D 198 EHHQGTGKWCYIPMYDFHICISDALBEGITHSCLTEFPQDNRLYDWDNTITIPVHPRO 257
Qy 432 EFSRLNMYTLLSKRKLMPVONKVEDWTPRPFTVOGIVRGLKYEALIQFILOQAS 491
D 258 EFSRLNMYTLLSKRKLMPVONKVEDWTPRPFTVOGIVRGLKYEALIQFILOQAS 491
Qy 492 KNLNLEMDKLTINKKIIDPVCAHNTAVLKDQRIFFLTNGPRE-EPPVRLIPRHKREG 550
D 318 KQDNTIEASLESICREBLNENAPRAMVIDPVKLV-TEVPQGESEMTVPNHPKPE 375
Qy 551 AGKATTPANRIMLDYAD-AAAIK-----GEEVTIMDQNA-IYKIKYESVITEL 601
D 376 MGSREVPFSGEIMWRADFPREANQYKRLVWGKEVRLR---NAVYIAERVEKA----- 428
Qy 602 VGEHLLEGSVKT-----TKLK--ITWLADIEELVPLSLVEFDYLISK 642
D 429 -----EGNITITICTYDADTLISKPADGRKVGVIHWVS-VAHALPIEIRLYDLFESVP 481
Qy 643 KLEDEDFLDNLNPTCRREIPALGDANRNKIRGGIIOLEKGYVRCAPFIRSKPV 700
D 482 NPGAEDFLSYINPESLVIKOGFAEPSLKAAVAKAFQEREGYCLDRHSTAEKPV 539

RESULT 9
glutamine-tRNA ligase (EC 6.1.1.18) [validated] - Escherichia coli (strain K-12)
N/Alternate names: glutaminyl-tRNA synthetase
C/Species: Escherichia coli
C/Date: 13-Jun-1983 #sequence revision 05-Dec-1997 #text_change 03-Jun-2002
C/Accession: G64802; A92346; S03376; A01190; A31223
R/Bletter: F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
.A.; Rose, D.U.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: G64802
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-554 <BLAT>
A/Cross-references: GB:M10187; GB:U00096; NID:G1786888; PIDN:AACT3774.1; PID:G1786895;
R/Experimental source: strain K-12, substrain MG1655
R/Yamao, F.; Imokuchi, H.; Cheung, A.; Ozeki, H.; Söll, D.
J. Biol. Chem. 257, 11639-11643, 1982
A/Title: Escherichia coli glutaminyl-tRNA synthetase.
A/Reference number: A92346; MUID:83007336; PMID:6288695
A/Accession: A92346
A/Molecule type: DNA
A/Residues: 2-548, 'GR', 551 <YAM>
A/Cross-references: GB:M10187; NID:G146174; PIDN:AAA23884.1; PID:G146175
R/Uemura, H.; Conley, J.; Yamao, F.; Rogers, J.; Soell, D.
Protein Seq. Data Anal. 1, 479-485, 1988
A/Title: Escherichia coli glutaminyl-tRNA synthetase: a single amino acid replacement re
A/Reference number: S03376; MUID:89113343; PMID:2464170
A/Accession: S03376
A/Molecule type: DNA
A/Residues: 2-554 <UEM>

R/Hoben, P.; Royal, N.; Cheung, A.; Yamao, F.; Biemann, K.; Söll, D.
J. Biol. Chem. 257, 11644-11650, 1982
A/Title: Escherichia coli glutaminyl-tRNA synthetase. II. Characterization of the glnS g
A/Reference number: A92347; MUID:83007237; PMID:6798844
A/Contents: annotation; confirmation of amino and carboxyl ends by amino acid analysis;
C/Genetics:
A/Gene: glnS
A/Map position: 15 min
A/Function:
C/Description: EC 6.1.1.18 [validated, MUID:89113343]
A/Pathway: protein biosynthesis
C/Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology
C/Keywords: aminocyl-tRNA synthetase; ATP; ligase; protein biosynthesis
F/28-310/Domain: glutamine-tRNA ligase homology <BDL>

Query Match 20.2%; Score 753; DB 1; Length 554;
Best Local Similarity 35.2%; Pred. No. 2e-42;
Matches 189; Conservative 91; Mismatches 205; Indels 52; Gaps 14;

Qy 199 DLPAKVGKVCVRPAPEPSGLYHGHAKALINKYFAERYOGRILVFPDDTPNPKESNEF 258
D 19 DLAGSKHTVTRPPPEPNGYLHGHAKSICLNFGIADYQGCNLRFPDDTPNPKEDIEY 78
Qy 259 VENLKDIETLGIKYDA-VYTSDFPPLKMEMASBLIKOGAKYIDDTPEQMRKER---- 313
D 79 VDSIKNDVEMLGFHWSGDIRYSSDFDQHAHYAELINKGLAYDELTPEDIREYRGLT 138
Qy 314 MDGIESRCRNNTVEENLSLWKEM-VNGTERGQCCVCRKGLDMODPNKSLRDPVYRCNTD 372
D 139 APGKNSPRDRSVEENLALFEKRAAGFPBEG-KACLRKAKIDMASPFYMRDPLVLRIFA 197
Qy 373 PPHRVGSKYKVPYTFDACPVDALGVTHALRSSEYHNDRAOYRRLIDOMGLR-RVEIY 431
D 198 EHHQGTGKWCYIPMYDFHICISDALBEGITHSCLTEFPQDNRLYDWDNTITIPVHPRO 257
Qy 432 EFSRLNMYTLLSKRKLMPVONKVEDWTPRPFTVOGIVRGLKYEALIQFILOQAS 491
D 258 EFSRLNMYTLLSKRKLMPVONKVEDWTPRPFTVOGIVRGLKYEALIQFILOQAS 491
Qy 492 KNLNLEMDKLTINKKIIDPVCAHNTAVLKDQRIFFLTNGPRE-EPPVRLIPRHKREG 551
D 318 KQDNTIEASLESICREBLNENAPRAMVIDPVKLV-TEVPQGESEMTVPNHPKPE 375
Qy 551 AGKATTPANRIMLDYAD-AAAIK-----GEEVTIMDQNA-IYKIKYESVITEL 602
D 376 MGSREVPFSGEIMWRADFPREANQYKRLVWGKEVRLR---NAVYIAERVEKA----- 427
Qy 602 VGEHLLEGSVKT-----TKLK--ITWLADIEELVPLSLVEFDYLISK 643
D 428 -----EGNITITICTYDADTLISKPADGRKVGVIHWVS-VAHALPIEIRLYDLFESVP 481
Qy 644 LEEDEDFLDNLNPTCRREIPALGDANRNKIRGGIIOLEKGYVRCAPFIRSKPV 700
D 482 NPGAEDFLSYINPESLVIKOGFAEPSLKAAVAKAFQEREGYCLDRHSTAEKPV 538

RESULT 10
glutamine tRNA synthetase [imported] - Escherichia coli (strain O157:H7, substrain R1MD
F90717
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C/Accession: F90717
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hatboro, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: F90717
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-554 <HAY>
A/Cross-references: GB:BA000007; PIDN:BA834133.1; PID:G13360168; GSPDB:GN00154
A/Experimental source: strain O157:H7, substrain R1MD 0509952

Db 19 DLASGKHTTVHTRFPPEPNGYLHIGHAKSICLNFGIAQDYKGGQCNLRFDDTNPVKEDIEY 78

QY 374 HHRVGSKKYKVPYDYDFACPFVDALGEVTHALRSSEYHNRNAQYYRILQDMGLR-RVEIYE 432

QY 374 HHRVGSKKYKVPYDYDFACPFVDALGEVTHALRSSEYHNRNAQYYRILQDMGLR-RVEIYE 432

Db 199 HHGSGNKKCIYPMWDFTHCISDALEGIHSLCTLEFQNRRLYDVLNINISIDCHPRYE 258
Qy 433 FSRLLMVTLLSKRKLMLFVONKXVEDWTPRPFTVQIVRGKVEALIOFILOOGASK 492
Db 259 FSRLLMVTLLSKRKLMLFVONKXVEDWTPRPFTVQIVRGKVEALIOFILOOGASK 318
Qy 493 NLNLMEMKMTINKKIIDPCARHTAVLKQORVIFTLTNGPEEPFVAILPRHKKFEAG 552
Db 319 QNNVEMMSLSCIRDDLNENHAPRAMAVLDPKVIENRAAGEEMLT-MPHHPNPMG 376
Qy 553 KKATTPANRIMLDYAD-AAALNK-----GEEVTLMDGNA-LYKEIKVE---SGVITE 600
Db 377 SRQVPDESEIYIDRADPEEANKQYKRLVLGKEVRLR---NAVYKAERVEKDAENGNTT 433
Qy 601 LVGELHLE-----GSVTKTKLKITWLDIEELVPLSLVEFDYLSKKKLEDEDFDLNL 654
Db 434 LYCYDAETLNKDPADGKRVKGVIMWVS-VAHALPAEIRLYDRLPFNVPAPAAEDFLSTI 492
Qy 655 NP---CTPREI--PALGDANMRNIRGEIIQLEKGYRCDAPIRSKPPVYL 702
Db 493 NPESLVIRQGFVEPSLADA-----VSDKTYQFEREGYFCADS--RYSRPGAL 537

RESULT 13

H82254
glutaminyl-tRNA synthetase VC0997 [imported] - Vibrio cholerae (strain N16961 serogroup
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: H82254
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Diegel, I.; Sellere, F.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A62035; MUID:2040683; PMID:10952301
A:Accession: H82254
A:Status: preliminary
A:Residues: 1-556 <HEI>
A:Molecule type: DNA
A:Cross-references: GB:AE004181; GB:AE003852; NID:99655454; PIDD:AAF94158.1; GSPDB:GN001
A:Experimental source: serogroup O1, strain N16961, biotype El Tor
C:Genetics:
A:Gene: VC0997
A:Map position: 1
C:Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology

Query Match 19.3% Score 721; DB 2; Length 556;
Best Local Similarity 34.3%; Pred. No. 2.8e-40;
Matches 185; Conservative 91; Mismatches 208; Indels 56; Gaps 13;

Qy 199 DLPGAQKVCVRPAPEBSGYLHGHAKAALLNKYPARYQRLIVRPDTPNPKSENEF 258
Db 19 DLADGKHTHTVRPEPENGYLHGHAKSICLNFGIADYQOCNLRFDTPNPKENLEY 78
Qy 259 VENLKDIETIGIKYDA-VTYSDFPKLMEASBLIKOGKAYIDDPTEQMRKER----- 313
Db 79 VESIKKDVTLGFDMSGECYSSDFDKLYEYALIELIKGLAYVDELPEQIRREYRGLT 138
Qy 314 MDGIESRCRNNVTENSLMKEMVNGTERGMQCVRGKLDMDQPNKSLRDPVYRCNTDP 373
Db 139 EPGKSPYRDRSVEENLALFEKMRAGEFAEGACIRAKIDMASSTIVMDPVLVVRAB 198
Qy 374 HHRVGSKKVPTTYDFACPFVDALEGVTHALRSSEYHNRNAQYRIILDMGLR-RVEIYE 432
Db 199 HHGSGNKKCIYPMWDFTHCISDALEGIHSLCTLEFQNRRLYDVLNINISIDCHPRYE 258
Qy 433 FSRLLMVTLLSKRKLMLFVONKXVEDWTPRPFTVQIVRGKVEALIOFILOOGASK 492
Db 259 FSRLLMVTLLSKRKLMLFVONKXVEDWTPRPFTVQIVRGKVEALIOFILOOGASK 318
Qy 493 NLNLMEMKMTINKKIIDPCARHTAVLKQORVIFTLTNGPEEPFVAILPRHKKFEAG 552

Db 319 QENNIEYSALES CIRDDLNNENAPRAMAVLDPVKLVLENFAGVETILLANHPNPKPEM-- 376
Qy 542 LPRHKKTEGAGKATTPANRIMLDYAD-AAALNK-----GEEVTLMDGNA-LYKEIKV 593
Db 377 -----GDREVPTRFEMIEREDFREERANKYKRLVKGKEVRLR--GAYVIAKRI 424
Qy 594 ---ESGVIT-----ELVGEHLGSSVTKTKLKITWLDIEELVPLSLVEFDYLSKK 642
Db 425 EKDEQGNITTFICGYDEPDLCKNPADG--RKVKVIMWVS-AEKGVAPRFPRLVRLFTVP 481
Qy 643 KLEDEDFDLNLCCTPREIPALGDANMRNIRGEIIQLEKGYRCDAPIRSKPPVYL 702
Db 482 NPGADNPAETINDESIVKQGVPEPSLVEAKPEFGYQFERMGYFCADN---KDSFGAL 538

RESULT 14

G64118
glutamine-tRNA ligase (EC 6.1.1.18) - Haemophilus influenzae (strain Rd KW20)
N:Alternate names: glutaminyl-tRNA synthetase
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
C:Accession: G64118
R:Flaischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: G64118
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-557 <TIGR>
A:Cross-references: GB:U32814; GB:L42023; NID:91574809; PIDD:AAC23001.1; PTD:91574816; T
A:Genetics:
C:Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology
C:Keywords: aminocacyl-tRNA synthetase; ligase; protein biosynthesis
F:36-319/Domain: glutamine-tRNA ligase homology <EBI>

Query Match 19.0% Score 710.5; DB 2; Length 557;
Best Local Similarity 34.5%; Pred. No. 1.4e-39;
Matches 182; Conservative 93; Mismatches 214; Indels 39; Gaps 13;

Qy 199 DLPGAQKVCVRPAPEBSGYLHGHAKAALLNKYPARYQRLIVRPDTPNPKSENEF 258
Db 27 DLASGKHSVHTRPEPENGYLHGHAKSICLNFGIADYQOCNLRFDTPNPKEDVEY 86
Qy 259 VENLKDIETIGIKYDA-VTYSDFPKLMEASBLIKOGKAYIDDPTEQMRKER----- 313
Db 87 VDSIKADVEMLGFKNKGEPKPRASDFDLVGYAVELIKGLAYVDELSPDEKREYRGLT 146
Qy 314 MDGIESRCRNNVTENSLMKEMVNGTERGMQCVRGKLDMDQPNKSLRDPVYRCNTDP 373
Db 147 EPGKSPYRDRSVEENLALFEKMRAGEFAEGACIRAKIDMASPFMVVRREVIYRIKFS 206
Qy 374 HHRVGSKKVPTTYDFACPFVDALEGVTHALRSSEYHNRNAQYRIILDMGLR-VBIY 431
Db 207 HHGSGNKKCIYPMWDFTHCISDALEGIHSLCTLEFQNRRLYDVLNINISIERPLQY 266
Qy 432 EFSRLNMVYTLISKRKLMLFVONKXVEDWTPRPFTVQIVRGKVEALIOFILOOGAS 491
Db 267 EFSRLNMGITLLSKRKLMLFVONKXVEDWTPRPFTVQIVRGKVEALIOFILOOGAS 326
Qy 492 KNLNLMEMKMTINKKIIDPCARHTAVLKQORVIFTLTNGPEEPFVAILPRHKKFEAG 551
Db 327 KQDNVVEYSALACIRDLNENAPRAMAVIDPVRVLI---ENFSEAVLTAPNHPNREL 383
Qy 552 GKATTPANRIMLDYAD-AAALNK-----GEEVTLMDGNA-LYKEIKVE---SGVIT 599
Db 384 GERQLPFEKELIIRADPREERANKQYKRLVLGKEVRLR---NAVYKAERVEKDAENGIT 440
Qy 600 -----ELVGEHLGSSVTKTKLKITWLDIEELVPLSLVEFDYLSKKKLEDEDFL 651

Db 441 TIFCTYDPELTGKNPAG--RKVKGVIHWSAVNN-HPAEFRILYRLFTVNPGEADIE 497
QY 652 DNLPCTRRREIPALGDAMRNIRKGEIIQLRKGYRCDAPIRSSKP 699
Db 498 SVLPNPSLVIKQGFVEQSLANAERKGYQFEREGYFCADS--KDSRP 542

Search completed: January 25, 2004, 17:14:35
Job time : 72 secs

RESULT 15

P83421

glutaminyl-tRNA synthetase PA1794 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: P83421

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: P83421

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-556 <STO>

A:Cross-references: GB:AE004605; GB:AE004091; NID:g9947771; PIDN:AA05183.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: glsB; PA1794

C:Superfamily: glutamine-tRNA ligase; glutamine-tRNA ligase homology

Query Match 18.9%; Score 704.5; DB 2; Length 556;
Best Local Similarity 35.3%; Pred. No. 3.5e-39;
Matches 191; Conservative 81; Mismatches 216; Indels 53; Gaps 17;

QY 197 EVDLPAGVGVCVFPADEPSGYLHGHAKALINKYFAERYQGRIVRFDITNPSKESN 256
Db 18 QADIDAGHAKIVTRFPPEPNGLYHGHAKSICLNFGLAQFAGDCHLRFDDTNPAKEDQ 77
QY 257 EFVENLKDITLIGIKYNA-VTYSDFPKLMEASLTKOGKAVIDTTPKEQKER-- 313
Db 78 EYIDALEADIMVIGQWGSVCYASNYFDQLHAWAVELIKAGKAFVCDLGPPEMREYRG 137
QY 314 --MDGIESRCRNNTVEENLSLMEWVNGTERGMQCCVAGKLDMDPNKSLRDPVYYRCNT 371
Db 138 LTERGRNSPYRDRSVEENLDFARKKAGEFPDGRASLPAKIDMGSPNNMLADPILYIRIH 197
QY 372 DPHRVGSKYVYPTYPDFVDALEGVTHALRSSEYHNRNAOYRILQDML-RRVEI 430
Db 198 AHHTQTDKWCICYSDPTFGSDALEGITHSICLTFEDHRPLYEMFLMLPVPAQPRQ 257
QY 431 YEFRLMNVYTLISKRLIMVQNKVVDWTDPRPTVQGIVRGLKVEALIQFLQGA 490
Db 258 YEFRLMNVYTLISRRKKQLVDEGHVSGWDDPRMSTLSGYRRGYTPESIRNFCMTGV 317
QY 491 SKNLNMEWMDLMTINKKIIDVPCARHTAVLKDOGVIFLTNGPEEPFVRI-LPRHKKFE 549
Db 318 NRAGGVVDIGMLEPSIRHDLATAPRAMCVLPLKV--ITNYPEGVENLELPRHK-E 374
QY 550 GAGKATTFANRIMLDVADAAN-----KGEVTLMDGNAIVK--EIKVESGVI 598
Db 375 DMGVRVLPFGRELFIDAGDFEEVPAGYKRLIPGGEVRLR--GSYVIRADEALIKADGNI 432
QY 599 TEL-----VGEHLLEGSVKTKLITWLADIEELVPLSLVEFDYLI---SKKLEED 647
Db 433 VELRCSYDPDTLKG-NPEG--RKVGVYHNV-PAEGSVCEVRLLYDLFRSANPEKAEKG 488
QY 648 EDFLNLNP-----CTRRREIPALGDAMRNIRKGEIIQLRKGYRCDAPIRSSKP 699
Db 489 GSFIDNINADSLQVLAGC--RAEBSLGQANPE-----DRQFEREGYFVADIKDSRPGK 541
QY 700 V 700
Db 542 V 542

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2004, 16:58:16 ; Search time 39 Seconds
(without alignments)

862.156 Million cell updates/sec

Title: US-09-831-683B-10

Perfect score: 3734
Sequence: 1 MEALSFKSDSPISIIICAA.....SSKEPVLFALPDGRQASLS 715

Scoring table: BIOSUM62.

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1574.5	42.2	716 1	SYEC_SCHPO
2	1562	41.8	724 1	SYEC_YEAST
3	1498	40.1	1714 1	SYEP_DROME
4	1437	38.5	1440 1	SYEP_HUMAN
5	757.5	20.3	554 1	SYO_SALT
6	755.5	20.2	554 1	SYO_SALT
7	753	20.2	553 1	SYO_SALT
8	750	20.1	553 1	SYO_SALT
9	747	20.0	553 1	SYO_SALT
10	743.5	19.9	555 1	SYO_SALT
11	734.5	19.7	556 1	SYO_SALT
12	725	19.4	559 1	SYO_SALT
13	721	19.3	556 1	SYO_SALT
14	710.5	19.0	557 1	SYO_SALT
15	705.5	18.9	552 1	SYO_SALT
16	704.5	18.9	556 1	SYO_SALT
17	690	18.5	571 1	SYO_SALT
18	685	18.3	580 1	SYO_SALT
19	682	18.3	580 1	SYO_SALT
20	678	18.2	562 1	SYO_SALT
21	677.5	18.1	579 1	SYO_SALT
22	666.5	17.8	562 1	SYO_SALT
23	666.5	17.8	579 1	SYO_SALT
24	664.5	17.8	579 1	SYO_SALT
25	653	17.5	571 1	SYO_SALT
26	651.5	17.4	572 1	SYO_SALT
27	636.5	17.0	570 1	SYO_SALT
28	624	16.7	570 1	SYO_SALT
29	594	15.9	809 1	SYO_SALT
30	589.5	15.8	809 1	SYO_SALT
31	575	15.4	852 1	SYO_SALT
32	573	15.3	786 1	SYO_SALT
33	572	15.3	811 1	SYO_SALT

34	561.5	15.0	775 1	SYO_HUMAN	P47897 homo sapien
35	559	15.0	551 1	SYE_ARCFU	O29979 archaeoglob
36	553	14.8	778 1	SYO_DROME	O9Y105 drosophila
37	548	14.7	566 1	SYE_SULTO	O971d0 sulfolobus
38	540.5	14.5	571 1	SYE_METAC	O8T52 methanobact
39	537.5	14.4	553 1	SYE_METH	O26157 methanobact
40	530	14.2	552 1	SYE_METM	O50543 methanobact
41	529.5	14.2	571 1	SYE_METRA	O8Txb7 methanopyru
42	510.5	13.7	571 1	SYE_METMA	O8Pw52 methanobact
43	504	13.5	575 1	SYE_SULTO	P95968 sulfolobus
44	502.5	13.5	570 1	SYE_PYRAB	O8Zu33 pyrobaculum
45	480.5	12.9	548 1	SYE_THEAC	O9Yjms thermoplasma

ALIGNMENTS

RESULT 1	SYEC_SCHPO	STANDARD;	PRT;	716 AA.
ID	O13775;			
AC	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DE	Probable glutamyl-tRNA synthetase, cytoplasmic (EC 6.1.1.17)			
DE	(Glutamate--tRNA ligase) (GluRS).			
GN	SPAC17A5.15C.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972;			
RX	MEDLINE=21848401; PubMed=11859360;			
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,			
RA	Sgouros K., Peat N., Hayles J., Baker S., Basham D., Bowman S.,			
RA	Brooke K., Brown D., Brown S., Chillingworth T., Churcher C.M.,			
RA	Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,			
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,			
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,			
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,			
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,			
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,			
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,			
RA	Skellern J., Simmonds M., Squares R., Squares S., Stevens K.,			
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,			
RA	Woodward J., Voicakert G., Aert R., Robben J., Grymptre B.,			
RA	Welfjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,			
RA	Gabel C., Fuchs M., Fritze C., Holzer E., Holzer D., Hilbert H.,			
RA	Borzym K., Langer I., Beck A., Leirich H., Reinhardt R., Poll T.M.,			
RA	Bozzyk K., Zimmermann W., Medler S., Wambut R., Purnelle B.,			
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,			
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,			
RA	Lucas R., Rocher M., Galliard C., Tallada V.A., Garzon A., Thode G.,			
RA	Dada R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,			
RA	Dominquez A., Revuelta J.U., Moreno S., Armstrong J., Forsburg S.L.,			
RA	Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,			
RA	Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;			
RT	"The genome sequence of Schizosaccharomyces pombe.";			
RL	Nature 415:871-880(2002).			
CC	-1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +			
CC	diphosphate + L-glutamyL-tRNA(Glu).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).			
CC	-1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcements)			

or send an email to license@lsb-sib.ch.

CC -----

CC EMBL/298849; CAB1515.1; -

CC PIR/ T37830; T37830.

DR HSSP; P00962; 1GTR.

DR GenDB.SPombe; SPAC17A5.15c; -

DR InterPro; IPR0004526; Gltx arch.

DR InterPro; IPR000924; Glu_cRNA-synt_1c.

DR InterPro; IPR001412; cRNA-synt_1.

DR Pfam; PF00749; cRNA-synt_1c; 1.

DR Pfam; PF03950; cRNA-synt_1c; 1.

DR PRINTS; PR00987; TRNASYNTHGLU.

DR TIGRfam; TIGR00463; gltx arch. 1.

DR PROSITE; PS00178; AA_cRNA_LIGASE_1; 1.

DR Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding.

KW Aminoacyl-tRNA synthetase; "HIGH" REGION.

FT SITE 215 224

FT SITE 441 445

FT BINDING 444 444 ATP (BY SIMILARITY).

FT SEQUENCE 716 AA; 80749 MW; 2ACE0A35ED393227 CRC64;

Query Match 42.2%; Score 1574.5; DB 1; Length 716;
 Best Local Similarity 49.5%; Pred. No. 1.1e-101;
 Matches 319; Conservative 102; Mismatches 201; Indels 23; Gaps 10;

QY 70 GNDIEFGHVVWLEAYPTLSGS----EFENACLFVNGFLASRTFLVGHGTTIDIAVW 125
 DB 81 GENDSL--VESWETA-SALAGNNHFLSLAQDLHLMRSLFVGYSLTSADFSIW 137
 QY 126 SNLAGIGRWESLSKSKYQNLVRFWNSIDSEYKEALNVVAFAFGKGIKSPAPSLKE 185
 DB 138 GALKSNMNAAGAVR-TGQYFNLAIRYKQWDSQ--NAVETMEEFKAVANISK-----KQ 188
 QY 186 KVHDSKDSAPDEVDPGAKVGVCVFAPBPESGYLHIGAKAALINKYFAERYQRLIVR 245
 DB 189 K--SSGPNY-EIGPDAIDGKVTRFPPEPSGYLHIGAKAALINQYFANKYHGLIYR 244
 QY 246 FDDTNPSENEFVNLKDIETLGIKYDAVTYTSDFPKLMEASLTKQKATIDTTP 305
 DB 245 FDDTNPSENEFVNLKDIETLGIKYDAVTYTSDFPKLMEASLTKQKATIDTTP 304
 QY 306 KEOMKEMMDGIESRCNNNTVENLSLWKEMVNGTERGMOCCVCRKLMODPNKSLRPV 365
 DB 305 VETMHEHTEGIPSGHRDRPESLEIISMDKSGDVGLKNCIRAKISETENPKMRDPV 364
 QY 366 YYRCNTDPMHVRGSKYKYPYDPAFCFVDALGVTALRSSEYHNRNAQYRRILQDKL 425
 DB 365 IYRCNLDPHHRGTGKRAYPTDFACPIVDSLGVTHALRTTEYDRNPLVQMMIKAMNL 424
 QY 426 RRVETYESRLMVTLLSKRLMFMVQKKYEDTDRFPVQGIYVRGLKVEALLQFI 485
 DB 425 RKIHWEFSRNMFPVTLTSKRLTEIVDHGLWGMDDPRFPVRGVRRRGMTIELQOYI 484
 QY 486 LQOGASKULNMEWKLWTINKKIIDPYCARHTAVLKDORVFTLTNGPEEPFVILPRH 545
 DB 485 VSGGSKUILIDMTSPATNKIIDPVAFRHTAVESGDVAKATLVNCPAPAYADRRRH 544
 QY 546 KKFEGAGKATTFANRIMLDVAAAANKGEVTLMDGNALVKEIKYE-SGVITELVGE 604
 DB 545 KKNPELGKKSI FANEILLIEQDAQSFKQDEEVLTLMDGNAAVRIINDASGKVTSLKE 604
 QY 605 LHLBSGVTTKLTITWLADIELVPLSLVEFPYLLSKKLEDEDFLNDNCTREIPA 664
 DB 605 LHLDDPFKTEKVTWLTADTEKTPVDLVDFYLLTKKLEGEENYKDFLTPTQTEFHSV 664
 QY 665 LGDAMRNIRKGEIIOLEKGYRCDAPFIRSSKVVLPFAIPDGR 709
 DB 665 PADVGIKNLKKGDIIOVERKGIYDVDF--DGTQAVLFNIPDG 707

RESULT 2
 SYEC YEAST STANDARD; PRT; 724 AA.
 AC P46655;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Glutamy1-cRNA synthetase, cytoplasmic (EC 6.1.1.17) (Glutamate--cRNA ligase) (Glurs) (P85).

GN YGL245W OR G0583 OR HR8724.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RA Frantz J D., Gilbert W.;

RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.

RC STRAIN=S288c;

RA Vanderbol M., Durand P., Portetelle D., Hilger F.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 1-146 FROM N.A.

RC STRAIN=S288c / FY1679;

RA Colesac E., Maillier E., Neter P.;

RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + cRNA(Glu) = Amp + diphosphate + L-glutamyl-cRNA(Glu).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@lsb-sib.ch).

CC -----

CC EMBL/ U32265; AAA78905.1; -

CC EMBL/ Z49149; CAAB9009.1; -

CC EMBL/ Z72767; CAAG9694.1; -

CC EMBL/ X94357; CAA64142.1; -

DR PIR; S53934; S53934.

DR HSSP; P00962; 1GTR.

DR SGD; S0003214; YGL245W.

DR InterPro; IPR004526; Gltx arch.

DR InterPro; IPR000924; Glu_cRNA-synt_1c.

DR InterPro; IPR001412; cRNA-synt_1.

DR Pfam; PF00749; cRNA-synt_1c; 1.

DR Pfam; PF03950; cRNA-synt_1c; 1.

DR PRINTS; PR00987; TRNASYNTHGLU.

DR TIGRfam; TIGR00463; gltx arch. 1.

DR PROSITE; PS00178; AA_cRNA_LIGASE_1; 1.

DR Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding.

KW Aminoacyl-tRNA synthetase; "HIGH" REGION.

FT SITE 226 235

FT SITE 453 457

FT BINDING 456 456

FT CONFLICT 225 225 E -> D (IN REF. 1).

FT CONFLICT 489 489 V -> A (IN REF. 1).

FT CONFLICT 526 526 P -> S (IN REF. 1).

FT CONFLICT 562 562 V -> M (IN REF. 1).

FT CONFLICT 714 724 GKSYNKGAK -> VNLSTSMVQRNKHISNVTYLCYS

FT TSTP (IN REF. 1).

FT SEQUENCE 724 AA; 82662 MW; 34669BFB9CD41EB CRC64;

Query Match 41.8%; Score 1562; DB 1; Length 724;
 Best Local Similarity 45.7%; Pred. No. 8.1e-101;
 Matches 328; Conservative 121; Mismatches 212; Indels 56; Gaps 14;

QY 27 LTINHSIAAGAPTLPAGSGSLHGVN-----PIILYIARGASIASLSKNDIEFG-- 77
 DB 21 LTIN-----GKAPLVAAVELIARIIVNALPNSIAIKLVDDPKAPAKLDDATEDVFNKI 75
 QY 78 -----HVVEMLEIAPFL---SGSEFENACLFVDGFLASRTFLVGHGL--T 118


```

Db 76 TSKPAATFDNDKQVAKVMIAQKELVYKMFASQSLDTLSQNLRTFLIG-GLKYS 134
Oy 119 IADIAMNSNLGIGORWESLRSKKYONLVMPNSID-----SEKKEALNVAAFAVKR 173
Db 135 AADVACMGALNSNGM-CGSIINKKVDVANSRWYTLLENDPLFGEHADPLSKSLLEIKSA 193
Oy 174 GIGSPAPSLKKEVDSKDPAPNEVDLPQAKGVKCVAPAPESGYLHIGAKAALLNKY 233
Db 194 NVGK-----KKEHTKANF-----EIDLDPADKMGVEVTFPPPEPSYLIHIGAKAALLNOY 243
Oy 234 FAERYQGLIVRPDDTNPSSKSNFENVLLKDIEFLGIKYDAVYTSOYPPKLMEMASL 293
Db 244 FAQAVKGLHILRPDDTNPSSKEKEEFQDSILIEDLIGIKGDRITYSDDYFQGMVDYCVQM 303
Oy 294 IKQGAAYIDTTPKEKMRKERNDGIESRCRNNTVENENLSLW-KEMWNGTERKMGCCVRKRL 352
Db 304 IKDGAAYCDTTPTEKMRERMDGVASARRDRSVEENLRI FTEEMNGTEEGKKNCRVAKI 363
Oy 353 DMODPNKSLRDPVRYRCNTDPPHRRVGSKYKVPYDPACFPVDALEGVTHALRSSEYDR 412
Db 364 DYKALNKLRLRPVYIKRLCTFHHRGTSWKIYPTDFCVPIVDAIEGVTALRTIEYDR 423
Oy 413 NAOYRRIIDQNGLRARVEIYFSRLMVTYLLSKRKLWFVONKKVEDTDRPPTVQIV 472
Db 424 NAOYDMWLQALRLRVHILWDFARIVFVRLTLSSKRLQMMVWDKLVGNDPDRPFTVGR 483
Oy 473 RRGKLYELLQPILOQSGSKNLNMEMQKLTINKKIIDPCARFANLVKQOYVFTLTN 532
Db 484 RRGVTEELRNFVLSQSGSRVNIEMNLWAFNKKVIDPLAPRTALVNPVKHLEBSE 543
Oy 533 GPEEPFVILPRHKKFEGAGKATTFANRIMLDVADAANKGSEVTLMGNAIVKXIK 592
Db 544 APQPKIMKRRKHKKNPAVGEKVIYKYIDVDKDAVINVDEVTLMGNAIITK-K 602
Oy 593 VESGVITELVGLHLEGSVKTTLKITWADIIEELVPLSLVEFDYLLSKKLEDEDFLD 652
Db 603 NDDG---SMVAKLNLNEGDFKTKTKLTLWADTKDVPVADVDFHLIKDLLEBDESED 659
Oy 653 NLNCTREIRIPALGDANRNKIKGGIITOLEKGYKRCAPFRSSKPVVLAIPDGR 709
Db 660 FLTPQTEFHTADIADLVNKMKGIIQIFERKGYRLDA-LPKQKPYVFFTIIDPK 715

RESULT 3
STEP DROME STANDARD: PRT; 1714 AA.
ID STEP DROME
AC P28668; Q9VCF5;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bifunctional aminoacyl-tRNA synthetase [includes: Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase); Prolyl-tRNA synthetase (EC 6.1.1.15) (Proline--tRNA ligase)].
GN AATS-GLUPRO OR CGS394.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92097547; PubMed=1756734;
RA Cerini C., Kerjan P., Astier M., Gratecos D., Mirande M., Semeriva M.;
RT "A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA synthetase.";
RL EMBL J. 10:4267-4277(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Oregon-R;
RA Cerini C., Semeriva M., Gratecos D.;
RT "Evolution of the aminoacyl-tRNA synthetase family and the

```

```

RT organization of the Drosophila glutamyl-prolyl-tRNA synthetase gene.
RT Intron/exon structure of the gene, control of expression of the two
RT mRNAs; selective advantage of the multienzyme complex.";
RL Eur. J. Biochem. 244:176-185(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkely;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abil J.F., Agbavani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Baau A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Ferraz C., Fertiza S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostait N., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,
RA Jatala M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matlei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pectel J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -1- CATALYTIC ACTIVITY: ATP + L-proline + tRNA(Pro) = AMP +
CC diphosphate + L-prolyl-tRNA(Pro).
CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS
CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE
CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,
CC ARGINYL, AND ASPARYLYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY
CC PROTEINS, P18, P48 AND P43.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -1- SIMILARITY: Contains 6 WHEP-TRS domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; U74104; AAA28594.1; -
CC EMBL; U59923; AAC47469.1; -
CC EMBL; AE003745; AAF56211.1; -
CC PIR; S18644; S18644.

```

DR HSPB, P00962; 1GTR.
 DR FLYBae; FBgn005674; Acta-glupro.
 DR InterPro; IPR004526; GltX_arch.
 DR InterPro; IPR000924; GltX_cRNA-synt_1c.
 DR InterPro; IPR000406; GST_Cterm.
 DR InterPro; IPR004154; HGT anticonodon.
 DR InterPro; IPR004499; Pros_fam_1.
 DR InterPro; IPR002314; cRNA-synt_2b.
 DR InterPro; IPR001412; cRNA-synt_1.
 DR InterPro; IPR002316; cRNA-synt_pro.
 DR InterPro; IPR006195; cRNA_ligase_II.
 DR InterPro; IPR00738; WHEP-TRS.
 DR Pfam; PF00043; GST_C_1.
 DR Pfam; PF03123; HGT_anticonodon; 1.
 DR Pfam; PF00749; cRNA-synt_1c; 1.
 DR Pfam; PF03950; cRNA-synt_1c; 1.
 DR Pfam; PF00587; cRNA-synt_2b; 1.
 DR Pfam; PF00458; WHEP-TRS; 6.
 DR PRINTS; PR00987; TRNASYNTHGLU.
 DR PRINTS; PR01046; TRNASYNTHPRO.
 DR TIGRFAMs; TIGR00463; gltX_arch; 1.
 DR TIGRFAMs; TIGR00408; pros_fam_1; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
 DR PROSITE; PS00662; AA_TRNA_LIGASE_1; 1.
 DR PROSITE; PS00762; WHEP-TRS; 6.
 DR AMINOACYL-TRNA SYNTHETASE; protein biosynthesis; ligase; ATP-binding;
 DR MultiFunctional enzyme; Repeat.
 DR DOMAIN 170 754
 FT DOMAIN 755 800 WHEP-TRS 1.
 FT DOMAIN 827 872 WHEP-TRS 2.
 FT DOMAIN 901 946 WHEP-TRS 3.
 FT DOMAIN 980 1025 WHEP-TRS 4.
 FT DOMAIN 1055 1100 WHEP-TRS 5.
 FT DOMAIN 1129 1173 WHEP-TRS 6.
 FT DOMAIN 1174 1180 POLY-GLY.
 FT DOMAIN 1207 1714 PROLY-TRNA SYNTHETASE.
 FT SITE 209 220 "HIGH" REGION.
 FT SITE 438 442 "KMSKS" REGION.
 FT BINDING 441 441 ATP (BY SIMILARITY).
 FT CONFLICT 102 106 TSPPLP -> DKSIA (IN REF. 3).
 FT CONFLICT 233 234 VC -> AF (IN REF. 3).
 FT CONFLICT 341 345 NTACA -> KYCVR (IN REF. 3).
 FT CONFLICT 583 583 K -> R (IN REF. 3).
 FT CONFLICT 692 692 L -> A (IN REF. 3).
 FT CONFLICT 753 753 T -> S (IN REF. 3).
 FT CONFLICT 802 802 P -> T (IN REF. 3).
 FT CONFLICT 873 873 P -> T (IN REF. 3).
 FT CONFLICT 887 887 G -> V (IN REF. 3).
 FT CONFLICT 1201 1201 P -> PA (IN REF. 3).
 FT CONFLICT 1461 1461 MISSING (IN REF. 3).
 FT CONFLICT 1587 1587 G -> V (IN REF. 3).
 SQ SEQUENCE 1714 AA; 189197 MW; 6FE8C58045E48A8C CRC64;

Query Match 40.1%; Score 1498; DB 1; Length 1714;
 Best Local Similarity 44.2%; Pred. No. 8,1e-96;
 Matches 321; Conservative 126; Mismatches 247; Indels 32; Gaps 13;

QY 1 MEALSFKSDPPIIICAKLVG-LPLTINHSIAGSAPTLQFASGSL--HGVPPI 56
 DB 1 MSIKKALNNPPIGLATAHLINGTVPEIYMSKEES--LQPPDRRLVCHSNNDVL 57
 QY 57 LYIARGASIASISGNDIEFGHVEWLEVAFTLGSSEFNACLFVDGFLASRTFLVGHG 116
 DB 58 RALAAADPYKLYGTALERTQIDHWLSPLT--CEDISWALSPLTSPPLPVLYLVANK 115
 QY 117 LTIAIAWMSNLAGIGGQWESIRKSKYONLVWNSIDSEYKELNENVAFAVGRIG 176
 DB 116 LTIAADFAFNEM--HSRFEFLAAGKIPQHVQWDLTAQ--PLIQVLOSLPDARVK 170
 QY 177 KSPASLSE---KVHSDKDSAPEDVLPCAKYGVKCVFAPPSGYLHGHAKALLNKY 233
 DB 171 RSPGSSKQGTAKTERKO-EKGFVDLPQAEKGVVFRPPASGYLHGHAKALLNKY 229

QY 234 FAERYGRLYRFDNTNPKSENEFENLKDIEITLGIKYDAVYTSDYEPKLMEMAST 293
 DB 230 YALVCQGLINRFPDTPAKETVEFENVILLDDLOLKPVEFHTSYFDMIMDYCRL 289
 QY 294 IKQKAVYIDTPKQMKERMDGIESRCNNVTVENISLWKEMVNGTERGQCCVRGLD 353
 DB 290 IKESKAVYDTPPQOMLERQVRVESANRSVSKNLSLMEWVKGSGKQNTCAAKID 349
 QY 354 MODENSLRDPVYRCNTDPPHVRGSKYKVPYVDFACPFYDALEGVTHALRSSEYDRN 413
 DB 350 MSSNGCWDRDPTIRYCKNEPFRGTGYKVPYVDFACPIYDALENTHTLRTTEYHNRD 409
 QY 414 AQYRIIDMDGLREVEIYEPSERLMMVYTLISKRLTLPVQKKVEDMDPEFPVQIVR 473
 DB 410 DQFWFDALKLRPYVWSYRLMTVTIVLSKRLTWFSGLVDGDDPFPVGRILR 469
 QY 474 RGLKEALIOFLIOGASKNLNLMEMDKLTINKKIIDPVCARHTAVLKQDQVIFTLNG 533
 DB 470 RGMTEGKEFTIIOGSSKSVFPMNDKIMAFNKKVIDPIAPRYTALKEKRVLVNAGA 529
 QY 534 PESPFRVRI-LPRHKKFEGAGKATTPANRWLDYADAANKGEVTLMDGNAIVKEI- 591
 DB 530 KVE---RQVSVHFKDSSLGKQVTLAPRIYIDYVDAEALKEGNATFINNGNILLIKVN 586
 QY 592 KVESGVITELVGEHLLEGSVTTTLKLTWLA---DIEELVPLSVEPDIYLSKKLEDE 648
 DB 587 KDSAGNTSVDAALNLENKQFKTLKTLAVEDDPSAAYPFCVFPDNIISKAVLGDE 646
 QY 649 DFLNINPCTREIRIPALGDANMRNIRKGEITQERKGYVRCADAFIRSS-----KPVVL 702
 DB 647 DFKPFGIKTRDVPMLQDPELKKCKKGDIIQLQRGPFKVDVAVLPPSGVTNVPSPVL 706
 QY 703 FAIPDG 708
 DB 707 FSPIDG 712

RESULT 4
 SYEP HUMAN STANDARD; PRT; 1440 AA.
 AC ID P07814;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Bifunctional aminoacyl-tRNA synthetase (includes: Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--cRNA ligase); Prolyl-tRNA synthetase (EC 6.1.1.15) (Proline--cRNA ligase)).
 GN EPRS OR QPRS OR GLNS OR PARS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91107633; PubMed=1988429;
 RA Felt R., Knippers R.;
 RT "The primary structure of human glutamyl-tRNA synthetase. A highly conserved core, amino acid repeat regions, and homologies with translation elongation factors.";
 RL J. Biol. Chem. 266:1448-1455(1991).
 RN [2]
 RP PRELIMINARY SEQUENCE OF 96-887 FROM N.A.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=88262551; PubMed=3290852;
 RA Thomas P., Felt R., Schray B., Kunze N., Knippers R.;
 RT "The core region of human glutamyl-tRNA synthetase homologues with the Escherichia coli and yeast enzymes.";
 RL Nucleic Acids Res. 16:5391-5406(1988).
 CC -I- CATALYTIC ACTIVITY: ATP + L-glutamate + cRNA (Glu) = AMP + dihydrophosphate + L-glutamyl-tRNA (Glu).
 CC -I- CATALYTIC ACTIVITY: ATP + L-proline + cRNA (Pro) = AMP + dihydrophosphate + L-prolyl-tRNA (Pro).
 CC -I- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS

```

CC COMPRESSED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE. THE
CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,
CC ARGINYL, AND ASPARTYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY
CC PROTEINS, P18, P48 AND P43.
CC
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -1- SIMILARITY: Contains 3 WHEP-TRS domains.
CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A GLUTAMINYL-TRNA
CC SYNTHETASE.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A NUMBER OF
CC SEQUENCE ERRORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce
CC or send an email to license@isb-stb.ch).
CC
CC DR EMBL; X54326; CAA38224.1; -.
CC DR EMBL; X07466; CAA30354.1; ALT_SEQ.
CC DR PIR; A38663; SYHUOT.
CC DR PDB; 1PYU; 31-DEC-02.
CC DR Genew; HGNC:3418; EPRS.
CC DR MIM; 138295; -.
CC DR GO; GO:0005737; C:cytoplasm; TAS.
CC DR GO; GO:0005625; C:soluble fraction; TAS.
CC DR GO; GO:0006461; P:protein complex assembly; TAS.
CC DR InterPro; IPR004526; GLX arch.
CC DR InterPro; IPR000924; Glu tRNA-synt_1c.
CC DR InterPro; IPR004046; GST Cterm.
CC DR InterPro; IPR004154; HGF antileucodon.
CC DR InterPro; IPR004499; ProS fam I.
CC DR InterPro; IPR002314; tRNA-synt_2b.
CC DR InterPro; IPR001112; tRNA-synt_1.
CC DR InterPro; IPR002316; tRNA-synt_pro.
CC DR InterPro; IPR006195; tRNA_ligase_II.
CC DR InterPro; IPR000738; WHEP-TRS.
CC DR Pfam; PF000443; GST C_1.
CC DR Pfam; PF03129; HGT antileucodon.
CC DR Pfam; PF00749; tRNA-synt_1c_1.
CC DR Pfam; PF03950; tRNA-synt_1c_1.
CC DR Pfam; PF00587; tRNA-synt_2b_1.
CC DR Pfam; PF00458; WHEP-TRS_3.
CC DR PRINTS; PR00987; TRNASYNTHGU.
CC DR PRINTS; PR01046; TRNASYNTHPRO.
CC DR TIGRFAMs; TIGR00463; glx arch_1.
CC DR TIGRFAMs; TIGR00408; proS fam I_1.
CC DR PROSITE; PS00178; AA tRNA LIGASE I; 1.
CC DR PROSITE; PS00862; AA tRNA LIGASE II; 1.
CC DR PROSITE; PS00762; WHEP-TRS_3.
CC DR Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
CC Multi-functional enzyme; Repeat; 3D-structure.
CC FT DOMAIN 92 687
CC FT DOMAIN 688 884
CC FT DOMAIN 935 1440
CC FT SITE 132 142
CC FT SITE 360 364
CC FT BINDING 363 363
CC FT DOMAIN 688 733
CC FT DOMAIN 761 806
CC FT DOMAIN 839 884
CC FT DOMAIN 887 919
CC SEQUENCE 1440 AA; 163026 MW; C4E185A0A41C204 CRC64;
CC
CC Query Match 38.5%; Score 1437; DB 1; Length 1440;
CC Best Local Similarity 43.5%; Pred. No. 1,1e-91;
CC Matches 284; Conservative 133; Mismatches 198; Indels 38; Gaps 8;

```

ID	SYQ_SALTI	STANDARD	PRT	554 AA.
Db	1	MEHTTIDHMLBESATLWSSCDSBFTSTINELNHCLSLRTYLGNLSLADLCWAATLIGNA	60	
Qy	133	QRMWESLRKSKCYONLVRWVFNISIDSEYKALNEVVAFAVKGKIGKSPAPSLIKEVHDSKD	1920	
Db	61	AMQEQKQKKAPVHVKMFQFLEAQ--QAFQSV-----GTMKDVSTTKA	1020	
Qy	193	PSAPE-----VDLPKAVKVCVVRAPAPESGLIHLGAKKALLNKYPAERQGLIIV	2444	
Db	103	RVAPBKKQDVGRFVELPGAEKGVKVRPPPEASGLIHLGAKKALLNHQVQVNFKGLIM	1620	
Qy	245	RFDOTPSKSNBPFVNLLKDIETGICYDAVTYSYDPFPLTMAESLSLKOGKAYIDDT	3044	
Db	163	RFDPTNPEKEDKEFEVYLIEDVAMLIHPDPTTISDFETIMKYAEKLIQBGKAYVDT	2222	
Qy	305	PKEOMKERMDIGIESRCRNVTVEENLSMKEMVNGTERGMOCVGRKLDMDPNKSLRDP	3644	
Db	223	PAGQMAERQRRIESHKRNPIERKQLQMWEMKKGSQSGHSCCLRAKIDMSNNQCMRDP	2820	
Qy	365	VYTRCNTDPIHVRGSKTKYPTPYDACPVDALBSGVTALRSSEYHDAQAQYRILODMG	4224	
Db	283	TLVRCKIOPHPRGNKXNYPTPYDACPVIDSIEGVTALRTTEYHNRDEQFVIIIEALG	3424	
Qy	425	LRRVELYEPSRLNMYTLLSKRKLMPQONKVKDEWMDPRPPTQGIYRSLKTEALIOF	4844	
Db	343	IRKPYIWEYSRLNLTNTVLSKRKLTFVNEGVLQDMDPRPPTVGRVGLRQMTVEGLKQF	4020	
Qy	485	ILQOGASKYLINLMDMDKMTITINKKIIDVVCARHTAVLKDQGVITLTITNGDEEPPVRIIPR	5444	
Db	403	IAAQSSRSRVNMENDKIWAFAFKYVIDVAARYVALLKKEVILPNNVEAQEE--MKRVAK	4664	
Qy	545	HKKFEGAGKCAATTANRILWLDYADAALNKGEVTLMDMGNAIYKEI-KVESGYTELVG	6030	
Db	461	HPKNPPEVGLKPVWYSKVFIEGADLSEFSEGEWTFPIWGNLNTIKIHNKADGKILSLDA	5200	
Qy	604	ELHLEBSVATTKLKITWLAIDIELVPLSL--VEPPLYSKKKLEDEEDFLDNLNPCRRE	6614	
Db	521	KFPLENKDYKTKTKYTWLAETTHALPIPIVCYTEHLITKVLKIDEDPFOYVKNRKSKE	5800	
Qy	662	IPALGANNRNINRGEIITQLERKGYRCD-----APFIRSSKPVVLFAIPDG 708		
Db	581	ELMLGPPCLKDKLKGDIITQLQGRGFICDQPEPSPYSCNEARCVLIYIPDG 633		

[illegible]

SEQ_SALTY	RESULT 6
ID_SEQ_SALTY	STANDARD; PRT; 554 AA.
AC	Q8ZQX5;
DT	28-FEB-2003 (Rel. 41, Created)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
DE	(GlnRS).
GN	GINS OR STM0685.
OS	Salmonella typhimurium.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; Salmonella.
OX	NCBI_TaxID=602;
OX	1411

RP SEQUENCE FROM N.A.
RC STRAIN=L72 / SGSC1412 / ATCC 700720;
RX MEDLINE=21514948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Potwollik S., Ali T., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewel N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
L72.";
RL Nature 413:852-856(2001).
CC -I- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
CC diphosphate + L-glutamyl-tRNA(Gln).
CC -I- SUBUNIT: Monomer (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@isb-sib.ch).

DR	StyGeme; SG?????; glms.	
DR	HAMAP; MF_00126; -; 1.	
DR	InterPro; IPR0004514; GlnS.	
DR	InterPro; IPR000924; Glu_rRNA-synt_1c.	
DR	InterPro; IPR001412; rRNA-synt_1c.	
DR	Pfam; PF00749; rRNA-synt_1c; 1.	
DR	Pfam; PF03950; rRNA-synt_1c_C; 1.	
DR	PRINTS; PR00987; TRNASYNTHGLU.	
DR	TIGRFAMs; TIGR00440; glms; 1.	
DR	PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.	
KM	Aminoacyl-tRNA synthetase; Protein biosynthesis; Lysase; ATP-binding;	
KM	Complete proteome.	
FT	INIT_MET	0
FT	SITE	33
FT	SITE	267
FT	BINDING	270
FT	SEQUENCE	554 AA; 63406 MW; 8f47CFA9BA65BEAD CRC64;
Query Match	20.2%;	Score 755.5; DB 1; Length 554;
Best Local Similarity	35.3%;	Pred. No. 6.4e-45;
Matches 190; Conservative	93;	Mismatches 202; Indels 53; Gaps 15

QY 199 DLPGAKVGKVCVRAPPEPSGYIHHAKAALLINKYPAERYGRLLVREDTINPSKESNEF 258
 DB 18 DLASGKHITVTRFPPEPNNGYLIHGAKSICLNFIAADYQOCNLRFPDITPVKEDIEY 77
 QY 259 VENLLKDIETIGIKKDA-VTVSDPEPKLMEASBELIIOGKAYIDDTREQQRKR----- 313
 DB 78 VDSINNDVEWGLFPHSGDRIRSSDYFDQHHAYAVELINKGLAYDELPEQIREYRGLT 137
 QY 314 MDGIESRCNNVTENNLISLKMEN-VNGTERGMOCCVRGLMDQDNKSLRDVYRCNTD 372
 DB 138 APGKNSPFRDSSVEENLALFEKMRGGEPEG-KALCLRAKIDMASPIYMRDPVLYIRIFA 196
 QY 373 PPHRVGSKYKVPYTPDACPFDVDALEGVTHALRSSEYHDNRNAQYRRIIDMGLR-RVEIY 431
 DB 197 EHHQGNKMCITYPMWDFTHCISDALEGITHSLCTLEFQDNRLYQWVLDNITIPVHPROY 256
 QY 432 EFSRLNMYTLLSKRKLMLFVONKVEYEDTDRPFTVQGIYRGLKVEALIOFIIOGAS 491
 DB 257 EFSRLNLEYTWKSKRKLMLVTDKVEGWDPDRMPTISGLRRGYTAASIRBFCKRIGVT 316
 QY 492 KNLNLMEMDKLWTINKKIIDPVCARHTAVLKQRYIFLTNGPE-EPPVRIIPRHKKEG 550
 DB 317 KQDNTIEWASLESCTREDLNENAPAMAVIDPVKLV--IENYFQSESEMTVPHNPKPE 374
 QY 551 AGKKAATTPANRIWLDYAD-AAALNK-----GEEVTLMDWGN-IVKEIKVESGVITEL 601
 DB 375 MGSREVPSGEIWIADPFRKANQYKRLVWKGKVRLR---NAVYIAERKXA----- 427
 QY 602 VGEHLHESGKVT-----TKLK--ITWLADIIELVPLSLVEFDYLISK 642
 DB 428 -----ENITITICTYDADTLKSDPADGRKVKGYIHWVSAHAL-PIEIRLYDRHFGVP 480
 QY 643 KLEBEDFLDNLNPGCTRRREIPALGDANMNRKGIIOLEKGYRCAPFRSSKPV 700
 DB 481 NPGAEDFLSVINPESLVIKQGYGEPSLKAAYAGKAFPERGKGLDSTRATADKLV 538
 ID SYO_ECOLI STANDARD; PRT; 553 AA.
 AC P00962; Q59403;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine-tRNA ligase)
 GN GINS OR B0680.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 CX NCBI_TaxID=562;
 RN RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89113343; PubMed=2464170;
 RA Uemura H., Conley J., Yamao F., Rogers J., Soell D.G.;
 RT "Escherichia coli glutaminyl-tRNA synthetase: a single amino acid
 RT replacement relaxes tRNA specificity.";
 RL Protein Seq. Data Anal. 1:479-485(1988).
 RN RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=83007236; PubMed=6288695;
 RA Yamao F., Inokuchi H., Cheung A., Soell D.G.;
 RT "Escherichia coli glutaminyl-tRNA synthetase. I. Isolation and DNA
 RT sequence of the glns gene.";
 RL J. Biol. Chem. 257:11639-11643(1982).
 RN RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85051900; PubMed=6389180;
 RA Hoben P., Uemura H., Yamao F., Cheung A., Swanson R., Summer-Smith M.,
 RA Soell D.G.;
 RT "Misaminoacylation by glutaminyl-tRNA synthetase: relaxed specificity
 RT in wild-type and mutant enzymes.";

RL Ped. Proc. 43:2972-2976(1984).
 RN RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatcher F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirtpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=6905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN RN
 RP SEQUENCE OF 1-12.
 RC STRAIN=K12 / EMG2;
 RX MEDLINE=97443975; PubMed=9298646;
 RA Link A.J., Robinson K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 RN RN
 RP SEQUENCE OF 1-5, AND CHARACTERIZATION.
 RC STRAIN=K12;
 RX MEDLINE=83007237; PubMed=6749844;
 RA Hoben P., Royal N., Cheung A., Yamao F., Biemann K., Soell D.G.;
 RT "Escherichia coli glutaminyl-tRNA synthetase. II. Characterization of
 RT the glns gene product.";
 RL J. Biol. Chem. 257:11644-11650(1982).
 RN RN
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=90069584; PubMed=2479982;
 RA Rould M.A., Perona J.J., Soell D., Steltz T.A.;
 RT "Structure of E. coli glutaminyl-tRNA synthetase complexed with
 RT tRNA(Gln) and ATP at 2.8-A resolution.";
 RL Science 246:1135-1142(1989).
 RN RN
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=91312443; PubMed=1857417;
 RA Rould M.A., Perona J.J., Steltz T.A.;
 RT "Structural basis of anticodon loop recognition by glutaminyl-tRNA
 RT synthetase.";
 RL Nature 352:213-218(1991).
 RN RN
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE=96230743; PubMed=9562563;
 RA Rath V.L., Sullivan L.F., Beijer B., Sprcoat B.S., Steltz T.A.;
 RT "How glutaminyl-tRNA synthetase selects glutamine.";
 RL Structure 6:439-449(1998).
 RN RN
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=20318761; PubMed=10860750;
 RA Sherlin L.D., Bullock T.L., Newberry K.J., Lipman R.S., Hou Y.M.,
 RA Beijer B., Sprcoat B.S., Perona J.J.;
 RT "Influence of transfer RNA tertiary structure on aminoacylation
 RT efficiency by glutaminyl and cysteinyl-tRNA synthetases.";
 RL J. Mol. Biol. 299:431-446(2000).
 CC CC
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC -!- diphosphate + L-glutaminyl-tRNA(Gln).
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.emb-ebi.ch/announce/>
 CC or send an email to license@sb-ebi.ch).
 CC -----
 DR EMBL; V01575; CAA24894.1; -
 DR EMBL; J01617; AAA87980.1; -
 DR EMBL; M0187; AAA23884.1; -
 DR EMBL; AE000171; AAC73774.1; -
 DR EMBL; D90707; BAA35328.1; -
 DR EMBL; M16470; AAA69006.1; -
 DR EMBL; M16368; AAA69006.1; JOINED.
 DR PIR; G64802; SYECOT.
 DR PDB; 1GSG; 15-APR-92.
 DR PDB; 1GTR; 15-SEP-93.
 DR PDB; 1GTS; 15-SEP-93.
 DR PDB; 1QRS; 07-DEC-96.
 DR PDB; 1QRT; 07-DEC-96.
 DR PDB; 1QRU; 07-DEC-96.
 DR PDB; 1QTO; 27-MAY-98.
 DR PDB; 1EUQ; 04-JUN-00.
 DR PDB; 1EUD; 04-JUN-00.
 DR PDB; 1EXD; 03-AUG-00.
 DR PDB; 1NVL; 25-FEB-03.
 DR Phosbase; P00962; -
 DR SMSS-2DPAGE; P00962; COLI.
 DR EC02BASE; G061.0; 6TH EDITION.
 DR EcoGene; EG10390; glns.
 DR HAMAP; MF_00126; -; 1.
 DR InterPro; IPR004514; glns.
 DR InterPro; IPR000924; Gln tRNA-synt_1c.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c_C; 1.
 DR PRINTS; PR00987; TRNASYNTGLU.
 DR TIGRFAMs; TIGR00440; glns; 1.
 DR PROSITE; PS00178; AA tRNA_LIGASE I; 1.
 DR tRNA-cytoplasmic synthetase; protein biosynthesis; Ligase; ATP-binding;
 KW 3D-structure; Complete proteome.
 FT INIT MET 0
 FT SITE 33 "HIGH" REGION.
 FT BINDING 267 "KMSKS" REGION.
 FT BINDING 270 ATP (BY SIMILARITY).
 FT CONFLICT 548 553 WAKYGE -> GRK (IN REF. 2).
 FT HELIX 10 21
 FT TURN 22 22
 FT STRAND 28 31
 FT STRAND 40 40
 FT HELIX 41 56
 FT TURN 57 58
 FT STRAND 60 65
 FT HELIX 70 72
 FT HELIX 75 87
 FT TURN 88 89
 FT STRAND 97 98
 FT HELIX 99 102
 FT HELIX 103 115
 FT TURN 116 117
 FT STRAND 119 122
 FT HELIX 126 132
 FT TURN 133 133
 FT TURN 144 147
 FT HELIX 150 162
 FT TURN 163 163
 FT TURN 167 168
 FT STRAND 171 174
 FT TURN 178 179
 FT HELIX 183 185
 FT STRAND 189 193

FT STRAND 198 198
 FT TURN 199 201
 FT STRAND 202 202
 FT TURN 203 203
 FT STRAND 207 209
 FT HELIX 211 222
 FT TURN 223 223
 FT STRAND 226 230
 FT HELIX 231 233
 FT TURN 234 236
 FT HELIX 237 246
 FT STRAND 254 258
 FT STRAND 263 263
 FT TURN 264 265
 FT HELIX 270 278
 FT TURN 279 280
 FT TURN 285 286
 FT TURN 288 289
 FT STRAND 292 292
 FT HELIX 293 299

Query Match 20.2%; Score 753; DB 1; Length 553;
 Best Local Similarity 35.2%; Pred. No. 9.5e-45;
 Matches 189; Conservative 91; Mismatches 205; Indels 52; Gaps 14;

QY 199 DLPGAKYKVCVAPRPPSGYLHIGAKALLNKYFAERVOGRVIVFPDDTNSKESNEF 258
 DB 18 DLASGKHTYHTPPPPNGYLHIGAKSLCLNFGIAQDYKQCNLFDFTDPKEDIEY 77
 QY 259 VENLKDIEFTLGIKYDA-VYTSDDYPPKLMEMASLTKQKAYIDDPKEQMRKER---- 313
 DB 78 VESTKIDVEMLGRHMSGNVYSSDYPDQLAYALIELINKGLAYDELTPQDINEYRGLT 137
 QY 314 MDGIESRCNNVYEENLSLWKEV-VNGTERGMOCCVRGKLDMDQPNKSLRDPVYRQNTD 372
 DB 138 QPGKNSPYRDSYVENLALPEKRRAGGFEEG-KACLRAKIDMASPFIVMRDPVLYRIKFA 196
 QY 373 PHRVSGKIVYPPYDPACFPVDALBEGVTALASSEVHDNAQYRRLQDMGLR-RVEIY 431
 DB 197 EHHQTKMGCIYPPYDTHGICISDALBEGITSLCTLEFQDNRRLDWVLDNITTPVHPRQY 256
 QY 432 EFSRLNVTLLSKRLKLMFVONKVEVDTPRPPTVQGIYVRGLKVEALIQFLQOGAS 491
 DB 257 EFSRLNLEIYVMSGRKLNLTVDKVGMDPPKPTISGLRRGYTAASIREPKRGVT 316
 QY 492 KNLNLMEMDKLMTINKKIIDPVCAHTAVLKDORVIFTLTNGPEEPFVRILPRHKKFEGA 551
 DB 317 KQDVTIEMASLESCTREDLNEMNAPRAMAVIDPVGLVIENYQGESE--MVTMPNHPNPREM 374
 QY 552 GKATTFANRIWLDYD-AAALNK-----GEVTLMDGNA-IYKEIKYESVITELV 602
 DB 375 GSRQVPFSGEIVTIDRADFREBANQYRLVYGEKVRIR--NAVVIATEREKDA----- 426
 QY 603 GELHLEGSVKT-----TKLK--ITWLADIEELVPLVEFDYLSKKK 643
 DB 427 -----EGVITTFCTYADTLISKDPADGRKVKYIHVVSAAHL-PVIRIRYDLFVQVN 480
 QY 644 LEEDEDFLDNLNPCTRREIPALGDANNRNIKRGEIIQLERKYRCAPFIRSSKPV 700
 DB 481 PGAADDFLSVINPESLVYKQGFAPSLKDNVAGKAFQFEREGYCLDSRHSTAEKPV 537

RESULT 8
 SYQ_ECOL6
 ID SYQ_ECOL6 STANDARD; PRT; 553 AA.
 AC QERTM4;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase) (GlnRS).
 GN GINS OR C0766.
 OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=217992;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=06:1H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesech P.,
 RA Raeko D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC diphosphate + L-glutamyl-tRNA(Gln).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AE016751; AAN79239.1; -.
 DR HAMAP, MF_00126; -; 1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c; 1.
 DR TIGRFAMs; TIGR00440; glnS; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
 KW Complete proteome.
 RN INIT_MET 0 BY SIMILARITY.
 FT SITE 33 43 "HIGH" REGION.
 FT SITE 267 271 "KMSK" REGION.
 FT BINDING 270 270 ATP (By SIMILARITY).
 FT SEQUENCE 553 AA; 63362 MW; FE89470BA128D8F9 CRC64;
 SO
 Query Match 20.1%; Score 750; DB 1; Length 553;
 Best Local Similarity 35.0%; Pred. No. 1.5e-44;
 Matches 188; Conservative 92; Mismatches 205; Indels 52; Gaps 14;
 QY 199 DLPGAQKVCVAPAPSPSGYLHGAARALANKYFAERYQRLIVRPDTPNSKSNF 258
 DB 18 DLASGKHTTVHTRFPPENGYLHGAHSICLNFGIADQYKQCNLRFDFTNPKEDIEY 77
 DB 78 VDSIKNDVEMWGFHSGNVRYSDFDQLHAYAILINKGLAYDELTPPEQIREYRGLT 137
 QY 259 VENLKDLETGIGKDA-VYTSDFPKLMEASLIQKRAYIDDTREKQKRRK- 313
 DB 138 OPGKSPKRDSSVENLALPEKMTGSGEBG-KACLRKIDIASPFYWRDPLVLRIFA 196
 DB 373 PPHRVSRYKYVPTDFACPFVDALEGVTHALRSSEYHNRNAQYRIIDQMLR-RVEIY 431
 DB 197 EHHQGNKWCICYPMWDFTHCISDALEGITHSLCTLEFDQNRRLYMWVLDNITIPHPROY 256
 QY 432 EFSRLNMYTTLISRKRLMFQNKVDEWMTDRPFTVOGIYRGLKVEALIOFIHQGAS 491
 DB 257 EFSRLNMYTTLISRKRLMFQNKVDEWMTDRPFTVOGIYRGLKVEALIOFIHQGAS 491
 DB 492 KNLNIMEDYKLTWIKKTIIDPYCAHRTAVLKOQRIYFLTLNGEPEPFRVILPRHKKEFGA 551
 DB 317 KQDNTIEVASLSCREDLNEAPAPAAVVIDPVKVIENYQEGE-RVWTEPHNHNKREM 374
 QY 552 GKKAITPAMRIWLDYAD-AAAIK-----GEEVTLMDWGA-IVKEIKYESSGIVTELV 602
 DB 375 GSRQVPFGEGLWIDRADPREBANQKRLVIGKEVRLR---NAVYIKAREYKDA----- 426

QY 603 GELHLEGSVKT-----TKLK-ITWLDIEELVPLSVERDYLSKKK 643
 DB 427 -----EGNITITLFCYDADTSLKOPADGRKXGVIHWMSAAHAL-PVEIRLYDRLPSPVN 480
 QY 644 LEEDEFLDNLNICTRREIPALGDNMNIRKGEIILQERKGYNCAPFIRSSRPV 700
 DB 481 FGADDFLVSINPESLVIKQGFAPSPSLKDVAAGKAFQEREGYFCLDSRHSTAEKPV 537
 RESULT 9
 ID SYQ_ECO57 STANDARD; PRT; 553 AA.
 AC O8X98;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutamyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
 DE (GlnRS).
 GN GlnS OR 20827 OR EGS0710.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=83334;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RX Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RX Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RX Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RX Grobeck E.J., Davis N.W., Lim A., Dimatanta E.T., Potamoustis K.,
 RX Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RX Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533 (2001).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RX Hayashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
 RX Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
 RX Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RX Kunara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22 (2001).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC diphosphate + L-glutamyl-tRNA(Gln).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AE005246; AAG55002.1; -.
 DR EMBL, AP002552; BAB34133.1; -.
 DR PIR, F85567; F85567.
 DR PIR, F90717; F90717.
 DR HAMAP, MF_00126; -; 1.
 DR InterPro; IPR004514; GlnS.
 DR InterPro; IPR000924; Gln_sRNA-synt_1c.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c; 1.
 DR PRINTS; PR00987; TRNASYNTGLU.
 DR TIGRFAMs; TIGR00440; glnS; 1.

DR PROSITE: PS00178; AA TRNA LIGASE I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KM Complete proteome.
 FT INIT MET 0 BY SIMILARITY.
 FT SITE 33 43 "HIGH" REGION.
 FT SITE 267 271 "KMSKS" REGION.
 FT BINDING 270 270 ATP (BY SIMILARITY).
 SQ SEQUENCE 553 AA; 63378 MW; A49EFIDE8A742567 CRC64;
 Query Match 20.0%; Score 747; DB 1; Length 553;
 Best Local Similarity 34.8%; Pred. No. 2.5e-44;
 Matches 187; Conservative 93; Mismatches 205; Indels 52; Gaps 14;

QY 199 DLPGAQKVCVRPAPEPSGYLHGHAKAALLNKYFAERYGRLIVRPDDTNPSEKSENF 258
 DB DLASGKHTVHTRPPEPENGYLHGHAKSICLNFGIADYQGCNLRPDDTNPVEDIEY 77
 QY 259 VENLKDIETLGIKYDA-VVYTSDFPKLMEASLHKGKAYIDDTPEQWRKER---- 313
 DB VDSIKNDVEMWLGFRKSGNVRYSDFDQLHAYALILINKGLAYDELTPREQIREYRGLT 137
 QY 314 MDGIESRCNNVTVENLSLWKEM-VNGTERGQCVCVRGLDMQDNKSLRDPVYRCNTD 372
 DB QPGKSPYRDRSVEENLALFEKMRGTGPEEG-KACLRAKIDWASPFIVRDPVLYRIKEA 196
 QY 373 PHHVGSKYKYPTYPDFACPFVDALSGYTHALRSSEYHDNRNAQYRIIDQMLR-RVEIY 431
 DB 197 EHHGTGNCMCYIPWDFTHCISDALEGITHSLCTLEFQDNRLYQWLVNDITIPVHPROY 256
 QY 432 EFSRLNMYTLLSKRKLMPYONKVEEDMTDPRFPYTGIVRGLKVEALLQFILOQAS 491
 DB 257 EFSRLNLYTWSKRLNLVTDKAVEGMDPRMPTISGLRRRGYTAASINEFCRIGVT 316
 QY 492 KNLNLMEMDKLMTINKKIIDPYCAHTAVLKDQRYFTLTNGPEEPFVRLIPRHKKFECA 551
 DB 317 KQDNTIEMASLESICREDLNENAPAMAVIDPVKLIENYQGESE-VWTMNMHNKPEM 374
 QY 552 GKKAITFPNRIWLYAD-AAALNK-----GEETLMDMGN-IVKEIKVESGYITELV 602
 DB 375 GSRQVPFGEIWIWDSDFREBANKQYKRLVIGKEVRL---NAVYIKAEKREKDA---- 426
 QY 603 GELHLEGSVKT-----TKLK--ITWLADIETPLSLVEEDVLYISKK 643
 DB 427 -----EGVITTFCTYDADTLISKDPRADGRKVGKGVHWSAHAL-PVEIRLYDLRFSVPN 480
 QY 644 LEEDEDFLDNINPCFTRREIPALGDANMNRKGEIIOERKGYRCDAFPIRSSKPV 700
 DB 481 PGADDFLSVINPESLVIKQGAEPSLKDAVAGKAFQPEREGYFCLDSRHSHTAEKPV 537

RESULT 10
 SYO_YERPE STANDARD; PRT; 555 AA.
 AC Q88DD9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutamine--cRNA synthetase (EC 6.1.1.18) (Glutamine--cRNA ligase)
 DE (GlnRS).
 GN GINS OR YPO2630 OR Y1205.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Tlball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebatina M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooke K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltham T., Hamlin N., Holtroyd S., Jagers K., Kariyasev A.V.,

RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIMS / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of Yersinia pestis KIM.";
 RL J. Bacteriol. 184:4601-4611(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + cRNA (Gln) = AMP +
 CC -1- dihydrophate + L-glutaminyl-cRNA (Gln).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

CC -----
 DR EMBL: AJ14153; CAC92873.1; "-"
 DR EMBL: AE013724; AAM84782.1; ALT_INIT.
 DR PIR: AB0321; AB0321.
 DR HAMAP: MF_00126; -; 1.
 DR InterPro: IPR004514; GlnS.
 DR InterPro: IPR000924; Gln cRNA-synt_1c.
 DR InterPro: IPR001412; cRNA-synt_1.
 DR Pfam: PF00749; cRNA-synt_1c; 1.
 DR Pfam: PF03950; cRNA-synt_1c; 1.
 DR PRINTS: PR00987; TRNASTYNTGHL.
 DR TIGRFAMs: TIGR00440; glnS; 1.
 DR TRNASTYNTGHL.
 DR PROSITE: PS00178; AA TRNA LIGASE I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KM Complete proteome.
 FT SITE 34 44 "HIGH" REGION.
 FT SITE 268 272 "KMSKS" REGION.
 FT BINDING 271 271 ATP (BY SIMILARITY).
 SQ SEQUENCE 555 AA; 63750 MW; AD3C68D6FB56C0B5 CRC64;
 Query Match 19.9%; Score 743.5; DB 1; Length 555;
 Best Local Similarity 36.8%; Pred. No. 4.4e-44;
 Matches 196; Conservative 80; Mismatches 214; Indels 43; Gaps 15;

QY 199 DLPGAQKVCVRPAPEPSGYLHGHAKAALLNKYFAERYGRLIVRPDDTNPSEKSENF 258
 DB DLASGKHTVHTRPPEPENGYLHGHAKSICLNFGIADYQGCNLRPDDTNPVEDIEY 77
 QY 259 VENLKDIETLGIKYDA-VVYTSDFPKLMEASLHKGKAYIDDTPEQWRKER---- 313
 DB VDSIKNDVEMWLGFRKSGNVRYSDFDQLHAYALILINKGLAYDELTPREQIREYRGLT 137
 QY 314 MDGIESRCNNVTVENLSLWKEM-VNGTERGQCVCVRGLDMQDNKSLRDPVYRCNTD 372
 DB QPGKSPYRDRSVEENLALFEKMRGTGPEEG-KACLRAKIDWASPFIVRDPVLYRIKEA 196
 QY 373 PHHVGSKYKYPTYPDFACPFVDALSGYTHALRSSEYHDNRNAQYRIIDQMLR-RVEIY 431
 DB 197 EHHGTGNCMCYIPWDFTHCISDALEGITHSLCTLEFQDNRLYQWLVNDITIPVHPROY 256
 QY 432 EFSRLNMYTLLSKRKLMPYONKVEEDMTDPRFPYTGIVRGLKVEALLQFILOQAS 491
 DB 257 EFSRLNLYTWSKRLNLVTDKAVEGMDPRMPTISGLRRRGYTAASINEFCRIGVT 316
 QY 492 KNLNLMEMDKLMTINKKIIDPYCAHTAVLKDQRYFTLTNGPEEPFVRLIPRHKKFECA 551
 DB 317 KQDNTIEMASLESICREDLNENAPAMAVIDPVKLIENYQGESE-VWTMNMHNKPEM 374
 QY 552 GKKAITFPNRIWLYAD-AAALNK-----GEETLMDMGN-IVKEIKVESGYITELV 602
 DB 375 GSRQVPFGEIWIWDSDFREBANKQYKRLVIGKEVRL---NAVYIKAEKREKDA---- 426
 QY 603 GELHLEGSVKT-----TKLK--ITWLADIETPLSLVEEDVLYISKK 643
 DB 427 -----EGVITTFCTYDADTLISKDPRADGRKVGKGVHWSAHAL-PVEIRLYDLRFSVPN 480
 QY 644 LEEDEDFLDNINPCFTRREIPALGDANMNRKGEIIOERKGYRCDAFPIRSSKPV 700
 DB 481 PGADDFLSVINPESLVIKQGAEPSLKDAVAGKAFQPEREGYFCLDSRHSHTAEKPV 537

374 HHRVSKYKYPTYPDFACPFVDALSGYTHALRSSEYHDNRNAQYRIIDQMLR-RVEIY 432
 DB HHGSKNKCIYPMWDFTHCISDALEGITHSLCTLEFQDNRLYQWLVNDITIPVHPROY 258
 QY 433 FSRILNMYTLLSKRKLMPYONKVEEDMTDPRFPYTGIVRGLKVEALLQFILOQAS 492
 DB 259 FSRILNLYTWSKRLNLVTDKAVEGMDPRMPTISGLRRRGYTAASINEFCRIGVT 316
 QY 493 NLNLMEMDKLMTINKKIIDPYCAHTAVLKDQRYFTLTNGPEEPFVRLIPRHKKFECA 552


```

Db      319 QDNVNMWMSLESCICDDLINEHAPRAMAVLDPKVIYIENRAGEEWLT--MPNHPNNPEWG 376
Qy      553 KKATTFANRIWLDYAD-AAAIK-----GEVTLMDMGNA-IYKEIKVE---SGVILE 600
Db      377 SRQVPSDSITIDRDPFREERANKQYKRLVLGKEVRLR--NAYVIAKEVEEDAGSNVTT 433
Qy      601 LVGEHLHE-----GSVTKTKLKITWLAIDIEELVPLSLVEFDYLLSKKLEDEDFLDNL 654
Db      434 LYCSYDAETLNKDPADGRKVKGVIMHVS-VAAHALPAEIRLRYDRFNVPNPAABEDFLSTI 492
Qy      655 NP---CTRREI--PALGDANMENIKRGETIIQERKGYRCDAPIRRSKPVVL 702
Db      493 NPESLIVRQGFVEPFLADA-----VSDKTYQFEREGYFCADS---RYSRPAAL 537

RESULT 11
STQ_VIBPA
ID      SYQ_VIBPA      STANDARD;      PRT;      556 AA.
AC      Q87G4;
DT      15-SEP-2003 (Rel. 42, Created)
DT      15-SEP-2003 (Rel. 42, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
DE      (GlnRS).
GN      Glns OR VP0832.
OS      Vibrio parahaemolyticus.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC      Vibrionaceae; Vibrrio.
OX      NCBI_TaxID=670;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=RIMD 2210633 / Serotype O3:K6;
RX      MEDLINE=22508454; PubMed=12620739;
RA      Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA      Iijima Y., Naito M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA      Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT      "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT      distinct from that of V. cholerae.";
RL      Lancet 361:743-749(2003).
CC      -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
CC      diphosphate + L-glutaminyl-tRNA(Gln).
CC      -1- SUBUNIT: Monomer (By similarity).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      HMAP; AP005075; BACS9095.1; -.
DR      HAMAP; MF_00126; -.
DR      PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW      Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW      Complete proteome.
FT      SITE      34      "HIGH" REGION.
FT      SITE      268     "RMSKS" REGION.
FT      BINDING    271     ATP (By similarity).
SQ      SEQUENCE    556 AA; 63969 MW; 355B526332146B44 CRC64;

Query Match      19.7%; Score 734.5; DB 1; Length 556;
Best Local Similarity 34.8%; Pred. No. 1.9e-43;
Matches 184; Conservative 91; Mismatches 201; Indels 53; Gaps 12;

Qy      199 DLPAKVGKVCYRFPSPSGYHIGAKAALNKTFAEYRQRLVRFPTDTPSPKSNSEF 258
Db      19  DLADKHTSVHTRFPSPGNYHIGAKSICLNFAGADPYQOCNMLRFDITPEKEDIEY 78
Qy      259 VENILKDIETIGIKVDA-VTYTSDYFPKLMEASLIGKRAYIDTDPREQRKER----- 313

```

```

Db      79 VESIKKDVNVLGFEDWGCVCYSSNFFDKLYEVAIELINKGLAYDELSPEQIREYRGTLK 138
Qy      314 MDGIESRCRRNTVEENLSLKMENVNCTERGQCCRGGLDMQDPKSLRDPYVYCNTPD 373
Db      139 EPGKSPYRDSVEENLALPEKTRAGEFEBGACIRAKIDGSSFWMDPVLRYVRPAT 198
Qy      374 HHRVGSKYKVPPTYFACPFVDALGCVTHALRSSSEYHRNAQYRILQDMGR-VWEIYE 432
Db      139 HQTGDKKICITPMYDFTHCISDALGITHSICITLFEFMRRLRYDVLNITITDCPHQYE 258
Qy      433 FSRLLMVTLLSKRLLMFVONKVEDWTDPRFPVQGIVRGRLVEALLIQFLIOGASK 492
Db      259 FSRLLMVEYVMSKRRLQNLVTEKLVGMDPMPFVSGIRRGFTPASIRECKEIGVTK 318
Qy      493 NLNLMENKLTINKKIIDPVCARITVLKQRYL-----FTLTNPEEPFARI 541
Db      319 QENMIEFSSLESCIRDDLINEAPRAMAVLDPKVIYIENRAGEVANTLTANHPNPEM-- 376
Qy      542 LPRHKKFPGAGKATTFANRIWLDYAD-AAAIK-----GEVTLMDMGNAIYKEIKV 593
Db      377 -----GERVPTFREWIREDPREERANKYKRLVLGKEVRLR--GAYVIAKEV 424
Qy      594 E---SGVIT-----ELVGEHLHESGVTYTKLKITWLAIDIEELVPLSLVEFDYLLSKK 642
Db      425 EKDAEGNITTTTYCTYDPEFLGKNPADG--RKVKGVIMHVS-ADKALPAEIRLRYDRFLTPV 481
Qy      643 KLEDEDFLDNLNPFCTRREIPALGDANMENIKRGETIIQERKGYRCDA 691
Db      482 NPAAADEPASTINTDSLVLVINGFVPSLASAEAGYQFERMGYFCADS 530

RESULT 12
STQ_PASMU
ID      SYQ_PASMU      STANDARD;      PRT;      559 AA.
AC      P57847;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
DE      (GlnRS).
GN      Glns OR PM0528.
OS      Pasteurella multocida.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC      Pasteurellaceae; Pasteurella.
OX      NCBI_TaxID=747;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=PM70;
RX      MEDLINE=21145866; PubMed=11248100;
RA      May B.J., Zhang Q., Li L.L., Faustian M.L., Whitlam T.S., Kapur V.;
RT      "Complete genomic sequence of Pasteurella multocida Pa70.";
RL      Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC      -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
CC      diphosphate + L-glutaminyl-tRNA(Gln).
CC      -1- SUBUNIT: Monomer (By similarity).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE006088; AAK02612.1; -.
DR      HSSP; P00962; IGTR.
DR      HAMAP; MF_00126; -.
DR      InterPro; IPR004514; GlnS.
DR      InterPro; IPR000924; Gln tRNA-synt_1c.
DR      InterPro; IPR001412; tRNA-synt_1.

```

DR Pfam; PF00749; tRNA-synt 1c; 1.
 DR Pfam; PF03950; tRNA-synt 1c; 1.
 DR PRINTS; PRO0987; TRNASYNTHGU.
 DR TIGRFAMs; TIGR00440; glns; 1.
 DR PROSITE; PS00178; AA TRNA_LIGASE I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 44 54 "HIGH" REGION.
 FT SITE 279 283 "KMSKS" REGION.
 FT BINDING 282 282 ATP (BY SIMILARITY).
 SQ SEQUENCE 559 AA; 64309 MW; C3DBC90B65F071D1 CRC64;

Query Match 19.4%; Score 725; DB 1; Length 559;
 Best Local Similarity 33.5%; Pred. No. 8.6e-43;
 Matches 189; Conservative 101; Mismatches 223; Indels 52; Gaps 15;

QY 162 LNEVVAEVGKRGKIGKSPAPSLKEKVDHSDKPSADEVLPGAKVVCVCPAPESGYLH 221
 DB 5 VENVVA-----ENTKRPNTNFRQID-----DLASGKHGVQTRPPEENGYLH 51
 QY 222 IGHAKAALLNKYFAERYQRLIVRPDPTNPSKESNEFEVNLKDIETLGIKYDA-VITYS 280
 DB 52 IGHAKSICLNFGLADYQGLCNLRDPDPTNPVEDVEYDSIKQVEMLGFKWGESPRYAS 111
 QY 281 DYFPLTMAESLIQGKAYIDDTFKEQMKR-----MDGISRCNNTVENSLSMKEM 336
 DB 112 DYFQDLGYVALELEKGLAYVDELSPKREYRGTLTEPGKNSPYRDSIEENLAFBKM 171
 QY 337 VNGTERGQCCVRGLDMQDPNKSRLDPVYVRCNTDPHHRVSKYKVPYDFACPFYDA 396
 DB 172 KNGEIAEGKACIRAKIDMASPTVWRDVIYRIKATHQIGDKKCIIPMDFTHCISDA 231
 QY 397 LEGVTHALRSSEYHDNRNAYRIIDQMLR--VEIYFSRLNMYTLLSKRKLMPYQN 454
 DB 232 IERTHSCTLEFQONRLYDVLNENISARLPHQYEFSLNLESTILTSKRLQLVNE 291
 QY 455 KKVEDWTPRPPTVQGIYRGLKVALIQLIQQASKNLMDKMTINKIIDVYC 514
 DB 292 GIVGQWNPWRPTISGLRRRGTPASLSRPFRCIRIVTKQDNNVPSALEACIRDLNENA 351
 QY 515 ARHTAVLDQVIFLVTNGPEEPFRILPRHKKPFAGKAKATTFNRIWLDVAD-AAAIN 573
 DB 352 PRANAIVPLKIVENFSGKE---MLTAPNHNDELGVRELPTRELYIDADPREAN 408
 QY 574 K-----GEEVTLMDGNA-IVKEIKYESGV---IT-----ELVGEHLSEGSVKT 614
 DB 409 KQYKRLVVGKEVRLR---NAVYIKAREVKEAENITTYCTYDPTLGKPNADG-RKV 463
 QY 615 KLTITWLADIEELVLSVEFDYLSKKKLEDEDFDLNLPCTREIRPALGDANMRKIK 674
 DB 464 KGVLOWVS-AEDYLPAERFOYGRLTVANPGAAEDIHQVLPDLSVIKQGVVEKSIANAQ 522
 QY 675 RGEITLERKGYRCDAPFIRSKP 699
 DB 523 PEKAYQFEREGYTCADS---KDSRP 544

RESULT 13
 SYQ_VIBCH STANDARD; PRT; 556 AA.
 AC 09KTA6;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 28-FEB-2003 (Rel. 41; Last annotation update)
 DE Glutamyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
 GN (GlnRS).
 OS Glns OR VC0997.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10955301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umeyam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tetselin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
 RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae";
 RL Nature 406:477-483(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC -1- SUBUNIT: Monomer (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE004181; AAF94158.1; -
 DR PIR; H82254; H82254.
 DR HSSP; P00962; IGTR.
 DR TIGR; VC0997; -
 DR HAMAP; MF_00126; -; 1.
 DR InterPro; IPR004514; GlnS.
 DR InterPro; IPR000924; Gln tRNA-synt 1c.
 DR InterPro; IPR001412; tRNA-synt 1.
 DR Pfam; PF00749; tRNA-synt 1c; 1.
 DR Pfam; PF03950; tRNA-synt 1c; 1.
 DR PRINTS; PRO0987; TRNASYNTHGU.
 DR TIGRFAMs; TIGR00440; glns; 1.
 DR PROSITE; PS00178; AA TRNA_LIGASE I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 34 44 "HIGH" REGION.
 FT SITE 268 272 "KMSKS" REGION.
 FT BINDING 271 271 ATP (BY SIMILARITY).
 SQ SEQUENCE 556 AA; 64089 MW; 8108C0366F0760F8 CRC64;

Query Match 19.3%; Score 721; DB 1; Length 556;
 Best Local Similarity 34.3%; Pred. No. 1.6e-42;
 Matches 185; Conservative 91; Mismatches 208; Indels 56; Gaps 13;

QY 199 DLPQAKYGVCPAPESGYLHIGHAKAALLNKYFAERYQRLIVRPDPTNPSKESNEF 258
 DB 19 DLADGKHTTVTRPPEENGYLHIGHAKSICLNFGLADYQGLCNLRDPDPTNPEENILEY 78
 QY 259 VENLKDIETLGIKYDA-VITYSDYFPLTMAESLIQGKAYIDDTFKEQMKR----- 313
 DB 79 VESIKQVTVLGFDMGSEGVCEYDFKLYVALELEIQGLAYVDELTPEDIREYRGLT 138
 QY 314 MDGISRCNNTVENSLSMKEMVNGTERGQCCVRGLDMQDPNKSRLDPVYVRCNTDP 373
 DB 139 EPQKSPYRDSVENENALFEKRAAGEAEGACIRAKIDMASPTVWRDVLVLRVRAE 198
 QY 374 HHRVSKKYVPYDFACPFYDALEGVTHALRSSEYHDNRNAYRIIDQMLR-RVEIYE 432
 DB 199 HHQGDKKWCICYPMVDFTHCISDALEGITHSICLTFQONRLYDVLNENITRTPCHPRQYE 258
 QY 433 FSLNMYTLLSKRKLMPVQNKYVEDTDRPPTVQGIYRGLKVALIQLIQQGSK 492
 DB 259 FSRNLBETVWSKSKRLQVLTEKLVGTGWDPRMPTISGLRRGTPSPAIRFCRIGVTK 318
 QY 493 NLNLMEMDKMTINKIIDVPCARHTAVLKQDQRY-----FLTNGSEEPFVRI 541
 DB 319 QENNIEYSALESCTRDLDLENAPPAAMVLDVVKLVENFAGYVETLTLANHPKPEH-- 376

QY 542 LPRHKEGAGKATTPANRIMLDYAD-AAAIK-----GEEVTLMDGNAIYKEIKV 593
 DB 377 -----GDREVPFTRMIEBEDFREANKKKYRLVIGKEVRLL--GAYVIKARI 424
 QY 594 ---ESGVT-----ELVGEHLHESVTKTKITWLADIEELVPLSLVEFDYLISK 642
 DB 425 EKDKQGNITTFCSYDPELTGKNPADG--RKVGVIHWVS--AEKGVPAEPFLYERLFTPV 481
 QY 643 KLEBEDDLMDLNPCTRREIPALGDANMRNIKRGEIIQLERKGYRCAPFIRSKPVVL 702
 DB 482 NPGAADNPAETINPESLVKGVYPSLVEAKPEFGVPERGWYFCADN---KSSPOL 538

RESULT 14
 SYO_HAEIN STANDARD; PRT; 557 AA.
 ID SYO_HAEIN STANDARD; PRT; 557 AA.
 AC P43831;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
 DE (GlnRS).
 GN GINS OR H11354.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Kersavage A.R., Bule C.J., Tomb J.-F., Dougherty R.A., Kirkness E.F.,
 McLennan K., Sutton G., Fitzhugh W., Fields C.A., Goeyne J.D.,
 Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
 Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
 Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U32814; AAC23001.1; -;
 DR PIR; G64118; G64118.
 DR HSSP; P00962; 1GTR.
 DR TIGR; H11354; -;
 DR HAMAP; MF_00126; -; 1.
 DR InterPro; IPR004514; GINS.
 DR InterPro; IPR000924; Gln-tRNA-synt_1c.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF03950; tRNA-synt_1c; 1.
 DR PRINTS; PRO0967; TRANSTHEDU.
 DR TIGRFAMs; TIGR00440; glns; 1.
 DR PROSITE; PS00178; AA_tRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
 KW Complete proteome.

FT SITE 42 52 "HIGH" REGION.
 FT SITE 277 281 "KMSKS" REGION.
 FT BINDING 280 280 ATP (BY SIMILARITY).
 SQ SEQUENCE 557 AA; 64078 MW; 4FF1452D57A29F12 CnC64;
 Query Match 19.0%; Score 710.5; DB 1; Length 557;
 Best local similarity 34.5%; Pred. No. 8.7e-42;
 Matches 182; Conservative 93; Mismatches 214; Indels 39; Gaps 13;
 QY 199 DLPAKGVKVCVRAPSPSGYLHGHAKALLNKYFAPRYOGRILVPRPDTPNPKESNEF 258
 DB 27 DLASGKHKSVTRPPEPNGLYHGHAKSLINGLAEYGLCNLRDTPNPKEDVEY 86
 QY 259 VENILKDIETLGIKYDA-VTYSDFPPLMMAESLTKOGKAYIDTPPEQMRKER----- 313
 DB 87 VDSIKADVEMGFPMKEGEPRIASDYFDALYGYAVBELIKGLAYVDELSPDMREYRGILT 146
 QY 314 MDGIESRCRNTVEENISLWKEMVNGTERGNQCCVRGLDMQDPNKSIRDVPYRCNTDP 373
 DB 147 EPGKNSPYDRRTIRENLALFEKMGGEPAEGKALRAKIDWASPFWVREPVYIRIKESS 206
 QY 374 HHRVSKTKYVPTDFACPFDALEGVTHALRSSEYHNRNQQYRIIDQGLRR--VEIY 431
 DB 207 HHQGDWKVCYPMYDFHICISDAIERITHSICLIEFDQNRRLYDVMLENISIERPLPHQY 266
 QY 432 EFSRLNMYVTLISRKLLMFQNKVDEWTPRPPTVOGIYRGLKVALIOFLIOGAS 491
 DB 267 EFSRLNLEGITLSRKLLIKLVNDEIVDQNDPRMPTISGLRRGYTPASLAEFCRIGVT 326
 QY 492 KNLIMEMDKLMTINKKIIDPVCARHTAVLKDQVIFLTNGPEEPFVRLIPRHKKEGA 551
 DB 327 KQDVVEYSALAEACIRDELNENAPRAVAVIDPVRVVI---ENFSEAVLTAPNHNRPDL 383
 QY 552 GKAKTTPANRIMLDYAD-AAAIK-----GEEVTLMDGNA-IYKEIKV---SGVIT 599
 DB 384 GERDLPFKELIYIRADFREBANQYKRLVIGKEVRLL--NAVYIAERVEKQANGBIT 440
 QY 600 -----ELVGEHLHESVTKTKITWLADIEELVPLSLVEFDYLISKLEBEDFL 651
 DB 441 TIFCTYDELTIGKNPADG--RKVGVIHWVSVAVN--HBAERLRLDRLTVNPGAEIDIE 497
 QY 652 DNLNPCTRREIPALGDANMRNIKRGEIIQLERKGYRCAPFIRSKP 699
 DB 498 SVLNPNSLVIRKQGVVEQSLNAAEAKQGFREGEVFCADS---KDSRP 542

RESULT 15
 SYO_CLOPE STANDARD; PRT; 552 AA.
 ID SYO_CLOPE STANDARD; PRT; 552 AA.
 AC O8XMP3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
 DE (GlnRS).
 GN GINS OR GLTX OR CPB0645.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1502;
 RX MEDLINE=21664373; PubMed=11792842;
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 RT flesh-eater.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamine + tRNA(Gln) = AMP +
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003187; BAB80351.1; -
DR HAMAP; MF_00126; -; 1.
DR InterPro; IPR004514; Glns.
DR InterPro; IPR000924; Glu_tRNA-synt_1c.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c; 1.
DR PRINTS; PR00987; TRNASYNTHTLU.
DR TIGRfam; TIGR00440; glns; 1.
DR PROSITE; PS00178; AA_tRNA_LIGASE_1; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
KW Complete proteome.
FT SITE 33 43 "HIGH" REGION.
FT SITE 266 270 "KMSKS" REGION.
FT BINDING 269 269 ATP (BY SIMILARITY).
SQ SEQUENCE 552 AA; 63933 MW; A349EA7211F66FID CRC64;

Query Match 18.9%; Score 705.5; DB 1; Length 552;
Best Local Similarity 33.5%; Pred. No. 1.9e-41;
Matches 177; Conservative 95; Mismatches 209; Indels 47; Gaps 12;

QY 199 DLPGAQVGVCPVAPAPSPGYLHGHAKAALINKYFAERYQRLIVRPDDTNPSSKESNEF 258
DB 18 DLDSKHDITITRFPPEPNNGYLIHGHAKSIVLNFELGKEFNGRTNLRFDDTNPTEKDETX 77
QY 259 VENLLKDIETIGIKYDAVTYSDVFPKLMEMASLIKQKAYIDDTPKKQMKER---M 314
DB 78 VESTIEDVHWLGNSELSHFASNYFDEMVKRALLIKKAKAVCDLTPEETIKYGTLT 137
QY 315 DGIESRCNNTVEENLSLMEKVNNGTERGMOCVAGKLDMDPNKSLRDPVYRCNTDPH 374
DB 138 PGKESPPYNNRSEIENLDLFEERRKGEFEDGSKVLAKIDMSSPNINFRDPIIRIAHASH 197
QY 375 HRVGSKYVYPTYPACFPVDALSEGVTALNSSEYHNRNAQYRILQDMGLRVEI--- 430
DB 198 HNTGDKMCIYPMYDFAHPLDEAIEGITHSICLIEFAD----HRPLYDMFVKECEMESVP 252
QY 431 --YESSRLNMYTLLSKRKLMLFVONKKVVEDTDRPFTVOGIVRGLKVEALLIOFILO 488
DB 253 ROIEFARLINTNTVNSKKRLQLVDEGIVDGMDPRMPTVAGLRRGYTPKSIKRFCKAI 312
QY 489 GASKNLIMENDKLTINKKIIDPVCAHTAVLKDQVYFTLTNGPE--EPFVRLIPRHK 547
DB 313 GVAKADSTVDSGMLHFIREDDQETAPRAMAVINPLKV--ITNYPEGESSEILEIENPK 370
QY 548 FEGAGKATTPANRIWLDYADA-----AINKGEVTLMDGNAIVK--EIKVESG 596
DB 371 DESAGKRAVTFSEREYIREDDEMNPCKYFRLFPGNEVRLK--GAYFVKCNEVIKDENG 428
QY 597 VITELVGLHLEGSVKT-----TKLKTWLAIDIBELVPLSLVEDYLISKKLE- 645
DB 429 EVT---EIHCTYDEPETSGETGFGTKVKVGTIHWV--DANNCTPAEFRLYEPLIIDCEEN 483
QY 646 EDEDFLDNLNPTCTRREIPALGDANNRIKGEIILQERKGYRCDAF 693
DB 484 EKHFLBQINPNSLTICKGFIEPSAKDAKPODKYQLFRHGYNVDPNF 531

Search completed: January 25, 2004, 16:59:08
Job time : 40 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 25, 2004, 16:58:19 ; Search time 196 Seconds
(without alignments)
941.365 Million cell updates/sec

Title: US-09-831-683B-10
Perfect score: 3734
Sequence: 1 MEALSPKSDSPISIIICAA.....SSKEVLFALPDGRQASLS 715

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPRMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_plant:*
10: sp_rodent:*
11: sp_virus:*
12: sp_vertebrate:*
13: sp_unclassified:*
14: sp_virus:*
15: sp_bacteriophage:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3384	90.6	713	10 Q8S5M6	Q8S5M6 oryza sativ
2	3196	85.6	715	10 Q9L1Z8	Q9L1Z8 oryza sativ
3	2657	71.2	719	10 O82462	O82462 arabidopsis
4	2644.5	70.8	728	10 O65253	O65253 arabidopsis
5	2141	57.3	429	10 O8H8J4	O8H8J4 oryza sativ
6	1529	40.9	1714	5 O85TJ3	O85TJ3 oryza sativ
7	1516	40.6	1022	11 O8CGC7	O8CGC7 mus musculus
8	1339	35.9	1149	5 Q23J15	Q23J15 caenorhabditis
9	1287.5	34.5	717	4 O8NAJ6	O8NAJ6 homo sapien
10	1276	34.2	863	5 O8IDK7	O8IDK7 plasmodium
11	1139	30.5	642	5 O8SSR4	O8SSR4 encephalito
12	750	20.1	554	16 O8FJW4	O8FJW4 escherichia
13	685.5	18.4	918	5 O8IE10	O8IE10 plasmodium
14	673.5	18.0	791	10 Q9LQ07	Q9LQ07 oryza sativ
15	669	17.9	556	16 O8EGZ6	O8EGZ6 shewanella
16	653	17.5	549	16 Q8D2R6	Q8D2R6 wigglewort

17	651.5	17.4	580	17 Q8U064	Q8U064 pyrococcus
18	645.5	17.3	795	10 Q8W4F3	Q8W4F3 arabidopsis
19	610	16.3	248	5 O15562	O15562 naseema locu
20	599	16.0	786	10 Q9PR14	Q9PR14 arabidopsis
21	566.5	15.2	606	11 Q8RIV9	Q8RIV9 mus musculus
22	566.5	15.2	775	11 Q8BU21	Q8BU21 mus musculus
23	566.5	15.2	775	11 Q8BML9	Q8BML9 mus musculus
24	562.5	15.1	612	4 Q96AW5	Q96AW5 homo sapien
25	561.5	15.0	608	4 Q9BUZ3	Q9BUZ3 homo sapien
26	561.5	15.0	612	4 Q9BUZ3	Q9BUZ3 homo sapien
27	548	14.7	566	17 Q97ID0	Q97ID0 sulfolobus
28	540.5	14.5	579	17 Q8TYS2	Q8TYS2 methanobac
29	529.5	14.2	571	17 Q8TXB7	Q8TXB7 methanopyru
30	526	14.1	697	5 Q8SR10	Q8SR10 encephalito
31	510.5	13.7	571	17 Q8PW52	Q8PW52 methanobac
32	502.5	13.5	570	17 Q8ZU33	Q8ZU33 pyrobaculum
33	492	13.2	250	5 O15563	O15563 trichomonas
34	482.5	12.9	305	5 O15561	O15561 naseema locu
35	464	12.4	548	17 Q979Q0	Q979Q0 thermoplasma
36	438	11.7	531	5 Q95W41	Q95W41 schistosoma
37	431	11.5	581	1 Q9P9C2	Q9P9C2 uncultured
38	414.5	11.1	394	16 Q8DFN1	Q8DFN1 vibrio vuln
39	381.5	10.2	240	11 Q8BTR1	Q8BTR1 mus musculus
40	381.5	10.2	586	17 Q9HQ11	Q9HQ11 halobacteri
41	319.5	8.6	168	11 Q9CRF9	Q9CRF9 mus musculus
42	288.5	7.7	147	16 Q8DFN0	Q8DFN0 vibrio vuln
43	277.5	7.4	481	16 Q8RB93	Q8RB93 thermoplasma
44	275.5	7.4	485	16 Q97KC9	Q97KC9 clostridium
45	261.5	7.0	489	2 Q8GDY9	Q8GDY9 heliobacilli

ALIGNMENTS

RESULT 1
Q8S5M6 PRELIMINARY; PRT; 713 AA.
ID Q8S5M6
AC Q8S5M6;
DT 01-JUN-2002 (TREMBL:rel. 21, Created)
DT 01-JUN-2002 (TREMBL:rel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBL:rel. 23, Last annotation update)
DE Putative glutamyl-CRNa synthetase.
GN OJ1003C07.13.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriopharyngaceae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA McCombie W.R., Spiegel L., de la Bastide M., Preston R., Ferraro K.,
RA Kuit K., Nascimento L., Zucavert T., Balija V., Bell M., Baker J.,
RA Santos L., Miller B., Katzenberger F., Muller S., King L., Yang C.,
RA Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
OJ1003C07, from chromosome 10, complete sequence."
RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC113335; AM08833.1; -.
DR EMBL; AC113335; AM08833.1; -.
DR Gramene; Q8S5M6; -.
DR InterPro; IPR004526; G1X arch.
DR InterPro; IPR00924; G1X arch.
DR InterPro; IPR00412; CRNa-synt_1c.
DR Pfam; PF00749; CRNa-synt_1c; 1.
DR Pfam; PF03950; CRNa-synt_1c; 1.
DR TIGRPFAM; TIGR00463; G1X arch; 1.
DR PROSITE; PS00178; AA TRNA_LIGASE_1; 1.
KM Aminoacyl-tRNA synthetase; 80887 MW; 3E7C2020430DE803 CRC64;
SQ SEQUENCE 713 AA; 80887 MW; 3E7C2020430DE803 CRC64;
Query Match 90.6%; Score 3384; DB 10; Length 713;
Best Local Similarity 89.1%; Pred. No. 2.5e-234;
Matches 637; Conservative 44; Mismatches 32; Indels 2; Gaps 1;

```

Qy 1 MEALSFESKQSPSPISIIICAAKLVGLPTLNHSLAAGSAPTLQFASGSLSHGVPNLTIIYA 60
Db 1 MEAKLSFQSDSPSPISIIISAAKAVAGVSLSDISLGAAGSAPVLCFSSGSLSRGINLTIEYA 60
Qy 61 RGASIASISGKNDIEFGHVEMLEAPTFILSGSEENCLFVDSGLASTRTFLVGHGLTIA 120
Db 61 Q-SPPSLHGRDALIESGHVEMLEAPTFILSGSEENCLFVDSGLASTRTFLVGHGLTIA 118
Qy 121 DIAVMSNLAGIGQWESLRKSKYQNLVRFWNSIDSEYKEALNEVVAAPVGRKIGKSPA 180
Db 119 DITWMSNLAGIGQWESLRKSKYQNLVRFWNSIDSDYKDTLNEVIAAYVGRIGKSPA 178
Qy 181 PSLKEXHDSKDPASAPVVDLPGAKVGVCFVAPFSPSGYLHIGAKAALLNKYFAERYQ 240
Db 179 PNLKEXKVDSDKPSAPVVDLPGAKVGVCFVAPFSPSGYLHIGAKAALLNKYFAERYQ 238
Qy 241 RLIVRFDDTNPSSKESNEFVENLKDIEFLGIKYDAVVTSDYFPPKLMEMASLIKQKAY 300
Db 239 RLIVRFDDTNPSSKESNEFVENLKDIEFLGIKYDAVVTSDYFPPKLMEMASLIKQKAY 298
Qy 301 IDDTPEKQMRKERMDSIESRCRNNTVEENLSLMEKENVGTERGMQCCVRGKLDMDPNKS 360
Db 299 VDDTPKEQMRBERNDGVSCKRNNTVEENLSLMEKENVGTERGMQCCVRGKLDMDPNKS 358
Qy 361 LRDPVYRCNTDPPHHRVGSKYKVPYDFPACFPVDALGVTALRSSEYHNRNAQYRIL 420
Db 359 LRDPVYRCNTDPPHHRVGSKYKVPYDFPACFPVDALGVTALRSSEYHNRNAQYRIL 418
Qy 421 QDMGRARVEIYEFSLNMYVTLISKRKLMFQNKVDMDPRPTVOGIVRBLKTEA 480
Db 419 QDMGRARVEIYEFSLNMYVTLISKRKLMFQNKVDMDPRPTVOGIVRBLKTEA 478
Qy 481 LIQFLIOGASKNLNMEMDKLMTINKKIIPVCARHTAVLKDQRYIFLTITNGPEEPYR 540
Db 479 LIQFLIOGASKNLNMEMDKLMTINKKIIPVCARHTAVLKDQRYIFLTITNGPEEPYR 538
Qy 541 ILPRHKKFEGAGKATTFANRIWLDYADAALNKGEVTLMDGNAIVKEIKVESGVITE 600
Db 539 ILPRHKKFEGAGKATTFANRIWLDYADAALNKGEVTLMDGNAIVKEIKVESGVITE 598
Qy 601 LVGEHLHESGVTKTKITWLTADIEELVPLSLVEFDYILSKKLEBEDDFLDNLNPCR 660
Db 599 LVGEHLHESGVTKTKITWLTADIEELVPLSLVEFDYILSKKLEBEDDFLDNLNPCR 658
Qy 661 EIPALGDANNRNIRKGEIIOLEKGYRCADAPFIRSSKPVVFAIPDGRQOASLS 715
Db 659 ETLALGDANNRNIRKGEIIOLEKGYRCADAPFIRSSKPVVFAIPDGRQOASLS 713

```

RESULT 2
Q9L128 PRELIMINARY; PRT; 715 AA.

AC Q9L128; 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-OCT-2002 (Tremblrel. 22, Last annotation update)
OS Similar to glutamyl-tRNA synthetase.
OC Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone: P063B08."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001081; BAA90375.1; -.
DR HSBP; P00962; IGTR.
DR Gramene; Q9L128; -.

```

DR InterPro; IPR004526; GltX arch.
DR InterPro; IPR000924; GltX RNA-synt_1c.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c; 1.
DR PRINTS; PR00987; TRNASYNTHGU.
DR TIGRFAMs; TIGR00463; GltX_arch; 1.
DR PROSITE; PS00178; AA_tRNA_Ligase_I; 1.
KW Aminoacyl-tRNA synthetase.
SQ SEQUENCE 715 AA; 81053 MW; 9B2E55B019B62D CRC64;

```

Query Match 85.6%; Score 3196; DB 10; Length 715;
Best Local Similarity 82.6%; Pred. No. 8; le-221;
Matches 590; Conservative 74; Mismatches 48; Indels 2; Gaps 2;

```

Qy 1 MEALSFESKQSPSPISIIICAAKLVGLPTLNHSLAAGSAPTLQFASGSLSHGVPNLTIIYA 60
Db 1 MEPLAFPDSPSPISIIICAAKAVAGVSLTLDKLSGASAPTLHSGSDPIHGCVNLTIIYA 60
Qy 61 RGASIASISGKNDIEFGHVEMLEAPTFILSGSEENCLFVDSGLASTRTFLVGHGLTIA 120
Db 61 RVASVTSFGQDAILQAAVYDQMDLYAPVILSGSEFBAACSTLDGYLASRTFLVGSISIA 120
Qy 121 DIAVMSNLAGIGQWESLRKSKYQNLVRFWNSIDSEYKEALNEVVAAPVGRKIGKSPA 180
Db 121 DIAVMSNLAGIGQWESLRKSKYQNLVRFWNSI-ADYSDTLDEVSAAYVGRIGKIGSPA 179
Qy 181 PSLKEXHDSK-DPSAPVVDLPGAKVGVCFVAPFSPSGYLHIGAKAALLNKYFAERYQ 239
Db 180 PSLKEXKPDSDQNSISIEVDLPGAKVGVCFVAPFSPSGYLHIGAKAALLNKYFAERYK 239
Qy 240 GRLIVRFDDTNPSSKESNEFVENLKDIEFLGIKYDAVVTSDYFPPKLMEMASLIKQKGA 299
Db 240 GRLIVRFDDTNPSSKESNEFVENLKDIEFLGIKYDAVVTSDYFPPKLMEMASLIKQKGA 299
Qy 300 YIDTPKEQMRKERMDSIESRCRNNTVEENLSLMEKENVGTERGMQCCVRGKLDMDPNK 359
Db 300 YVDDTPKEQMRBERNDGVSCKRNNTVEENLSLMEKENVGTERGMQCCVRGKLDMDPNK 359
Qy 360 SLRDPVYRCNTDPPHHRVGSKYKVPYDFPACFPVDALGVTALRSSEYHNRNAQYRIL 419
Db 360 SLRDPVYRCNTDPPHHRVGSKYKVPYDFPACFPVDALGVTALRSSEYHNRNAQYRIL 419
Qy 420 LQDMGLARVEIYEFSLNMYVTLISKRKLMFQNKVVEDMDPRPTVOGIVRBLKTEA 479
Db 420 LQDMGLARVEIYEFSLNMYVTLISKRKLMFQNKVVEDMDPRPTVOGIVRBLKTEA 479
Qy 480 ALIOPFLIOGASKNLNMEMDKLMTINKKIIPVCARHTAVLKDQRYIFLTITNGPEEPYR 539
Db 480 ALVQFLIOGASKNLNMEMDKLMTINKKIIPVCGHHTAALKDQRYIFLTITNGPEEPYR 539
Qy 540 RILPRHKKFEGAGKATTFANRIWLDYADAALNKGEVTLMDGNAIVKEIKVESGVITE 599
Db 540 RILPRHKKYEBAGKATTFNKNILLESADASVISIGEVTLMDGNAIVKEIKYONGIIT 599
Qy 600 ELVGEHLHESGVTKTKITWLTADIEELVPLSLVEFDYILSKKLEBEDDFLDNLNPCR 659
Db 600 ELVGEHLHESGVTKTKITWLTADIEELVPLSLVEFDYILSKKLEBEDDFLDNLNPCR 659
Qy 660 REIPALGDANNRNIRKGEIIOLEKGYRCADAPFIRSSKPVVFAIPDGRQOAS 713
Db 660 QEPALGDANNRNIRKGEIIOLEKGYRCADAPFIRSSKPVVFAIPDGRQOAS 713

```

RESULT 3
O82462 PRELIMINARY; PRT; 719 AA.
AC O82462; 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Glutamy1-tRNA synthetase.
 GN AT5G26707.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eustoides II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98438737; PubMed=9765600;
 RA Day I.S., Golovkin M., Reddy A.S.;
 RT "Cloning of the cDNA for glutamyl-tRNA synthetase from Arabidopsis
 thaliana.";
 RL Biochim. Biophys. Acta 1399:219-224 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shimizu P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shimizu P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF067773; AAC36469.1; -;
 DR EMBL: AY098592; AAC20443.1; -;
 DR EMBL: BT000248; AAN15567.1; -;
 DR HSSP: P00962; 1GTR.
 DR InterPro: IPR004526; Glx arch.
 DR InterPro: IPR000924; Glu tRNA-synt_1c.
 DR InterPro: IPR001412; tRNA-synt_1.
 DR Pfam: PF00749; tRNA-synt_1c; 1.
 DR Pfam: PF03950; tRNA-synt_1c; 1.
 DR PRINTS: PRO0987; TRNASYNTHGU.
 DR TIGRFAMs: TIGR00463; gltx arch; 1.
 DR PROSITE: PS00178; AA tRNA_LIGASE_I; 1.
 KW Aminocyl-tRNA synthetase.
 SQ SEQUENCE 719 AA; 81064 MW; 0778C243219DA24C CRC64;
 Query Match 71.2%; Score 2657; DB 10; Length 719;
 Best Local Similarity 69.0%; Pred. No. 4,2e-182;
 Matches 491; Conservative 106; Mismatches 107; Indels 8; Gaps 5;

DB 305 AYVDDTPREOMQKERMMDGIDSCRNHSEENI.KLMKEMIAESRGLOCCVRGKNMODPN 364
 QY KSLRDPVYRCNTDPPHVRVSGKYKVPYDPACFPVDALEGVTHALRSSEYHNRVAYR 418
 DB 359 KAMDPPVYRCNPNHSHRIGYKIPYDPACFPVDSLBGITHALRSSEYHNRVAYR 424
 QY ILDDWGLRVEIYEFSLNMYTLLSKRKLMPYQNKVDEMTDPRPPTVQGVARGIKV 478
 DB 425 VLEMGARQVQLYEFSLNMYTLLSKRKLMPYQNKVDEMTDPRPPTVQGVARGIKV 484
 QY 479 EALIQFIIOGASQULNMEMDKIMTINKTIIDPVCAHVAVLKDQVIFLTLNGPEEP 538
 DB 485 EALIQFIIOGASQULNMEMDKIMTINKTIIDPVCAHVAVLKDQVIFLTLNGPEEP 544
 QY VRIIPRHKPEGACKKATTFANRIMLDVADAANKGEVTLMDGNVIVEI-KVESGV 597
 DB 545 VRIIPRHKPEGACKKATTFANRIMLDVADAANKGEVTLMDGNVIVEI-KVESGV 604
 QY 598 ITTELVEGLHLEGSVYKTKLKITWMLADIEELVPLSLVEFDYLLISKKLEDEDFLDNLN 657
 DB 605 VTALSGVNLNQGVSXKTKLKITWMLADIEELVPLSLVEFDYLLISKKLEDEDFLDNLN 664
 QY 658 TRREIPALGDANMENIRKGEIIOLEKRGYRCADAPFIRSSRPVLPALPDGR 709
 DB 665 TKKETLALGDSNMRLKCGDIVIOLEKRGYRCADAPFIRSSRPVLPALPDGR 716
 RESULT 4
 ID 065253 PRELIMINARY; PRT; 728 AA.
 AC 065253;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, last annotation update)
 DB F21E10.12 protein.
 GN F21E10.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eustoides II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Davidson S., Rohlfing T., David M., O'Brian D.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Washu;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterson R.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Wilson R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF058914; AAC13597.1; -;
 DR HSSP: P00962; 1GTR.
 DR InterPro: IPR004526; Glx arch.
 DR InterPro: IPR000924; Glu tRNA-synt_1c.
 DR InterPro: IPR001412; tRNA-synt_1.
 DR Pfam: PF00749; tRNA-synt_1c; 1.
 DR Pfam: PF03950; tRNA-synt_1c; 1.
 DR PRINTS: PRO0987; TRNASYNTHGU.
 DR TIGRFAMs: TIGR00463; gltx arch; 1.
 DR PROSITE: PS00178; AA tRNA_LIGASE_I; 1.

SQ SEQUENCE 728 AA: 82066 MW: F7D42BAEAD5A5E93 CRC64;
 Query Match 70.8%; Score 2644.5; DB 10; Length 728;
 Best Local Similarity 68.2%; Pred. No. 3.4e-181;
 Matches 492; Conservative 105; Mismatches 107; Indels 17; Gaps 6;
 5 LSPFSDSPPIITICAAKVGPLPTINHLAASATLPDPAAGESLHGVPILYIARA- 63
 6 LSPFSPSPPLSVIYALSLASPVITIDSSAAATVPSFVSOGKRLNGATVLLRYGRSAK 65
 64 SIASLSGNDIEFGHV-----EWLEAYAPRLSGSEFENACLFDFGFLASRTPLVG 114
 66 KLDPFYGNNAFDSQSVSLICINMKIDENVDIASVSSSEFENACGRDKLIESSTFLVG 125
 115 HGLTIADIAVSNLAGIQWRWSLRSKKYQNLVWMSIDSEYKALNEVVAAPVGRG 174
 126 HSLSIADVAIVSALAGTQWRWSLRSKKYQSLVWFSIDSEYKALNEVVAAPVGRG 184
 175 IGKS-PASLSKRYH----DSKDPSPAEVDLPGAIVGKVCVRPAPEEGYLIHGAAL 229
 185 SGKPVAAKSKDSQAVKGDGDKGKPEVDLPBAEIGYKURFAPEPSGYLIHGAAL 244
 230 LNKYFAERYOGLIYRPDTPNPKSENEFVENLKDIEITIGKYAVYTSDPYPKLMM 289
 245 LNKYFAERYOGEVYVRFDPNPKSENEFVDNLVKDITGIGKYKVTYSDYFPELMD 304
 290 AESLIKQKAYIDTPEKQMRKERMDGIESRCNNVTEENSLMKEMVNGTERGMQCVR 349
 305 AEXLMREKAYVDTPREOMQKERMWDGIDSKCRNHSVEENLKLMEWLAGSERGLQCCVR 364
 350 GKLMDQPNKSLRDPVYRRCNTDPHHRVGSCKKYPTYPDFACPFDALBEGYTHALRS 409
 365 GKFNNQDPNKMARDVYVRCNPMSHRIGDKYKIPYDFACPFVDSLEGYTHALRSSEY 424
 410 HDNNAOYRIIDQMDGRLRVEIYEFSLNMVYLLSKRLMVFONKRYEDMTDPFPVQ 469
 425 HDNNAOYRIIDQMDGRLRVEIYEFSLNMVYLLSKRLMVFONKRYEDMTDPFPVQ 484
 470 GIVRGLKVEALIQFILOQASKNLNMWMDLMTINKKIIDPVCARHTAVLKQORVIF 529
 485 GIVRGLKVEALIQFILOQASKNLNMWMDLMTINKKIIDPVCARHTAVLKQORVIF 544
 530 LTNGBEPFVRLPPLPHKKFEGAGKATTPANRIWLDYDAALINKGEEVYTLMDQNAIYK 589
 545 LITDGPDEPFVPMIPGKKFEGAGKATTPANRIWLDYDAALINKGEEVYTLMDQNAIYK 604
 590 EI-KYESGVIETELVGLHLEGSVKTTKLKITWLADIEELVPLSLVEFVYLLSKKLEDE 648
 605 EITKDEBGRVTLASGLVNLQGSVKTKLKITWLADIEELVPLSLVEFVYLLSKKLEDE 664
 649 DFLDNLNCTRRREIPALGDANNRNIKGEIIOLEKRGYRCDAPIRSGKPVLPFAIPDG 708
 665 EVADVNNTKTKETLALDGSNMNRNKKCGVIOLEKRGYRCDAPIRSGKPVLPFAIPDG 724
 QY 709 R 709
 DB 725 R 725
 RESULT 5
 Q8H8J4 PRELIMINARY; PRT; 429 AA.
 AC Q8H8J4;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative RNA synthetase.
 GN OSJBA0083M08.1.
 OS Oryza sativa (japonica cultivar-group).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Eriocaulaceae; Oryzaceae; Oryza.
 NCBI_TaxID=39947;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RC Buehl C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
 RA Overton II L.L., Teichert T., Kim M.M., Bera J.J., Jin S.S.,
 RA Fadron D.W., Tallon L.J., Koo H., Ziemann U., Hejiao J., Blunt S.,
 RA Vanaken S.S., Redmiller S.B., Utechtack T.T., Feldblum T.V.,
 RA Yang Q.Q., Haas B.U., Suh B.B., Peterson J.J., Quackenbush J.,
 RA White O., Salzberg S.L., Frazer C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJBA0083M08 genomic sequence."
 DR Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 KW EMBL; AC096691; AAN08648.1;
 KW Aminoacyl-tRNA synthetase.
 SQ SEQUENCE 429 AA: 49743 MW: E1A8B223946622 CRC64;
 Query Match 57.3%; Score 2141; DB 10; Length 429;
 Best Local Similarity 52.1%; Pred. No. 2.2e-145;
 Matches 395; Conservative 25; Mismatches 9; Indels 0; Gaps 0;
 287 MEMAESLIKQKAYIDTPEKQMRKERMDGIESRCNNVTEENSLMKEMVNGTERGMQC 346
 1 MEMAESLIKQKAYIDTPEKQMRKERMDGIESRCNNVTEENSLMKEMVNGTERGMQC 60
 347 CVRGLMDQPNKSLRDPVYRRCNTDPHHRVGSCKKYPTYPDFACPFDALBEGYTHALRS 406
 61 CVRGLMDQPNKSLRDPVYRRCNTDPHHRVGSCKKYPTYPDFACPFDALBEGYTHALRS 120
 407 SEYHDNNAOYRIIDQMDGRLRVEIYEFSLNMVYLLSKRLMVFONKRYEDMTDPFP 466
 121 SEYHDNNAOYRIIDQMDGRLRVEIYEFSLNMVYLLSKRLMVFONKRYEDMTDPFP 180
 467 TVOGIVRGLKVEALIQFILOQASKNLNMWMDLMTINKKIIDPVCARHTAVLKQORV 526
 181 TVOGIVRGLKVEALIQFILOQASKNLNMWMDLMTINKKIIDPVCARHTAVLKQORV 240
 527 IETLTNGBEPFVRLPPLPHKKFEGAGKATTPANRIWLDYDAALINKGEEVYTLMDQNA 586
 241 IETLTNGBEPFVRLPPLPHKKFEGAGKATTPANRIWLDYDAALINKGEEVYTLMDQNA 300
 587 IVEIKYESGVIETELVGLHLEGSVKTTKLKITWLADIEELVPLSLVEFVYLLSKKLE 646
 301 IVEIKYESGVIETELVGLHLEGSVKTTKLKITWLADIEELVPLSLVEFVYLLSKKLE 360
 647 DEDFLDNLNCTRRREIPALGDANNRNIKGEIIOLEKRGYRCDAPIRSGKPVLPFAIP 706
 361 DEDFLDNLNCTRRREIPALGDANNRNIKGEIIOLEKRGYRCDAPIRSGKPVLPFAIP 420
 QY 707 DGRQOASLS 715
 DB 421 DGRQOASLS 429
 RESULT 6
 Q95TL3 PRELIMINARY; PRT; 1714 AA.
 AC Q95TL3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE LPA42739P
 GN AATS-GLUPRO OR CG5394.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Flise E., George R.,
 RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouenavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AY058703; AAL13932.1; -
 DR PLAYbase; FB00005674; Aats-glupro.
 DR InterPro; IPR001589; Actbind_actin.
 DR InterPro; IPR004526; GltX_arch.
 DR InterPro; IPR000924; GltX_arch.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR004154; HGT anticon.
 DR InterPro; IPR004499; Pros_fam_1.
 DR InterPro; IPR002314; tRNA-synt_2b.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR InterPro; IPR002316; tRNA-synt_1.
 DR InterPro; IPR006195; tRNA_Ligase_II.
 DR InterPro; IPR000738; WHEP-TRS.
 DR Pfam; PF00043; GST_C_1.
 DR Pfam; PF01223; HGT_anticon.
 DR Pfam; PF00749; tRNA-synt_1c_1.
 DR Pfam; PF03950; tRNA-synt_1c_C_1.
 DR Pfam; PF00587; tRNA-synt_2b_1.
 DR Pfam; PF00458; WHEP-TRS_6.
 DR PRINTS; PR00987; TRNASYNTHLU.
 DR PRINTS; PR01046; TRNASYNTHLU.
 DR TIGRFAMs; TIGR00463; gltX_arch; 1.
 DR TIGRFAMs; TIGR00408; pros_fam_1; 1.
 DR PROSITE; PS00178; AA tRNA_Ligase_1; 1.
 DR PROSITE; PS00862; AA tRNA_Ligase_II; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00762; WHEP-TRS; 6.
 DR SEQUENCE 1714 AA; 18941 MW; 3F8CF3DB128765A8 CRC64;

Query Match 40.9%; Score 1529; DB 5; Length 1714;
 Best Local Similarity 44.6%; Pred. No. 1.7e-100; Indels 32; Gaps 13;
 Matches 324; Conservative 129; Mismatches 241;

QY 1 MEALSPSKDPPSIICAAKLVG--LPITNHSIAAGSAPTLQFASGESL-HGVNPII 56
 DB 1 MSITLKNANNPISGLATAHLINGTVPEIIVMSKEETS---LQPPDRLLVCHSNNDVL 57
 QY 57 LYIRGASIASLGNIDIEFGHVVEMLEAFTPLSGSFENACLFVDFGLASRTFLVGHG 116
 DB 58 RALRAAPDYKLYGFTALERTQIDHMLSFSLT--CEDDISWALSLDLSIAPIVLYLVANK 115
 QY 117 LTIDIAVMSNLAGIGQWESLRSKSKYQNLVRWNSIDSEYKALNVAAFPVKRIG 176
 DB 116 LTIDAFALFENM--HSRYEFLAAKGIPOHVQWYDLTAAQ--PIQLQVLSLPEDAKVK 170
 QY 177 KSPAPSLKE--KYHDSKDPAPAEVDLPAGAKGVCFAPAPESGYLHIGAKAALLNKY 233
 DB 171 RSPSSKEGTAKTGERKQ-BGKPYDLPGLAMGKVVAFPPBASGYLHIGAKAALLNOY 229
 QY 234 FAERYQGLIYRPDDTNPESKESNEFENLLKDIETLGIKYDAVYTSDFPKLMEASL 293
 DB 230 YALFQGLTILIRFPDTPNPAKTEVEFENVITGLDEQLIKPDVFTHTSNYFDLMDYCVRL 289
 QY 294 IKQKAYIDTPPKQMKERBDGHSRCNNNTVENLSLMKEMVNGTERGMOCCRGKLD 353
 DB 290 IKESKAYVDTPPEQMKLERQORVESANRSNSVEKNLSLMEEMVXGSEKGYCVRAKID 349
 QY 354 MODNKSILRDPVYRCNTDPHHRVGSKYKVPYTFDAPCFVDALLEGVTHALRSSEYHNRN 413
 DB 350 MSSNGCKRDPITIRCKNEPRHPTGTAKYVFTPDPAFPIYDAIENVHTILRTTYHNRD 409
 QY 414 AQQYRILODMGLRVEIYFESRLNNVYLLSKRKLMEFVQNKKVEDWTPRFPYQGIIVR 473
 DB 410 DQFWFIDALKLRKPYIWSYSLNNTNTVLSKRKLTFVDSGLVDGMDPFRFPYVRIIR 469
 QY 474 RGLVEALIQITLLOGASRNINLMWDKLTINKKIIPVCAKRTAVLKQORVIFTLNG 533
 DB 470 RGMVTEGKEPTLIAGSSKSVFPMWDMKIMAFNKKVIDIPRIYALKEKRVINNVAGA 529
 QY 534 PEEFVRI-LPRHKKFEGAGKATTFARIMLDYADAALINKEVEVTLMGNAIVKRI- 591

DB 530 KVE---RIQVSHPKDESIGKTVLLGPRIYIDYDAEALKEGENATFINMGNIIRKVN 586
 QY 592 KVESGVITTELGEHLGSGVXTTKLTITWLA---DIEELVPLSLVEFDYLSKXKLEDE 648
 DB 587 KQASGNITSDAALNLENKPKFKTLKLTWAVEDDPSAYPPTFCYFNDITISKAVLGDE 646
 QY 649 DFLDNLPCTRRERIPALGDANMREIKGEIITLERKGYRCADAFIRSS-----KPVVL 702
 DB 647 DFKQFIGHKTRDEVPMIGDPELKKCKGDIITQLRGPFKVDVAVAPPSGYTNVPSPIVL 706
 QY 703 FALPDG 708
 DB 707 FSLPDG 712

RESULT 7

ID Q8CGC7 PRELIMINARY; PRT; 1022 AA.
 AC Q8CGC7;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Similar to glutamyl-prolyl-tRNA synthetase (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId:10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RA Strausberg R.;
 RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
 KW EMBL; BC040802; AAH40802.1; -
 FT Aminoacyl-tRNA synthetase.
 FT NON TER 1022
 SQ SEQUENCE 1022 AA; 114132 MW; B2EBD862D891AB61 CRC64;

Query Match 40.6%; Score 1516; DB 11; Length 1022;
 Best Local Similarity 42.9%; Pred. No. 6.6e-100; Indels 40; Gaps 10;
 Matches 310; Conservative 147; Mismatches 225;

QY 5 LSPSKDPPSIICAAKLVG--LPITNHSIAAGSAPTLQFASGESLHGVNPIIYARGAS 64
 DB 6 LTVAAGNPPEALLAVEHVKDVIS--SVBGRKNLIRVSTVAFDTNLSIRLARIAT 63
 QY 65 IASLSPKNDIEFGHVVEMLEAFTPLSGS- FENACLFVDFGLASRTFLVGHGLTADIA 123
 DB 64 TSGLYTNLMETHEIDHMLFESATKLSGCDGLTSAINEINHLCLRTLYVGNSTLADLC 123
 QY 124 VMSNLAGIGQWESLRSKSKYQNLVRWNSIDSEYKALNVAAFPVKRIGKSPASL 183
 DB 124 VMAITKSSAAMQEHKONKTLVHVRWFGFLFAO-----QAF--RSVG-----T 165
 QY 184 KEKHSKDPSPAP-----VDLPAGAKGVCFAPAPESGYLHIGAKAALLNKYFA 235
 DB 166 KMDVSGSRATYAPKQKQDVGFVLPGLAMGKVTYRPPBASGYLHIGAKAALLNGHYQ 225
 QY 236 ERYQGLIYRPDDTNPESKESNEFENLLKDIETLGIKYDAVYTSDFPKLMEASL 295
 DB 226 VNFQGLTILIRFPDTPNPAKTEVEFENVITGLDEQLIKPDQTYTSHDEFTIMKVAEKLIQ 285
 QY 296 QGKAYIDTPPEQMKERBDGHSRCNNNTVENLSLMKEMVNGTERGMOCCRGKLD 355
 DB 286 EGKAYVDTPPEQMKERBDGHSRCNNNTVENLSLMKEMVNGTERGMOCCRGKLD 345
 QY 356 DPNKSLRDPVYRCNTDPHHRVGSKYKVPYTFDAPCFVDALLEGVTHALRSSEYHNRN 415
 DB 346 SNGCKRDPITIRCKNEPRHPTGTAKYVFTPDPAFPIYDAIENVHTILRTTYHNRD 405
 QY 416 YRILQDMGLRVEIYFESRLNNVYLLSKRKLMEFVQNKKVEDWTPRFPYQGIIVR 475
 DB 406 FYWIELALGRKPYIWSYSLNNTNTVLSKRKLTFVDSGLVDGMDPFRFPYVRIIR 465

QY 476 LKVELLIOBILDOGASKNLMLNMEMDCLMTINKKIIDPVCARHNAVUKDOGVIFTLNNGPE 535
 Db 466 MIVEGKQCIITAAOGSSRSYVNNEWDKIMFNKVCIDPVARVYALLKKEVFPVNVLDAGE 555
 QY 536 EPPVRIILPHRKKTEGAGKATTPFANRIMLDYADAAAINKGEBYTLMDWGNAIKVEI-KYE 594
 Db 526 E--MKEVAPHKPNPDVGLKPRVWYSPIKVFEGADAEFFSEGEWTFINMGWINITKIHKNA 583
 QY 536 GSVITELVQELLEGSVKTKLKITLADIEBELVPS--LYVEDYLLSKKULEEDDFLD 652
 Db 556 DOKITSLDKMLNENDVKYKTKITITWLASTHALPIPAVCVITYEHLITKPVLEKDEDFQ 643
 QY 653 NLNPCRREIPALGDAMRNINIKGEIITOLEERKNYVCDAPIRFS-----KEVVLFAIP 706
 Db 644 YINKOSKHEMLMGDPCLDKLKKGDIITQIQRSGFFICDQYEVEVSPSCEHAPCILIYIP 703
 QY 707 DG 708
 Db 704 DG 705

```

RESULT 8
Q23315
ID Q23315 PRELIMINARY; PRT; 1149 AA.
AC Q23315;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE ZC434.5 protein.
GN ZC434.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
CX NCBI_Taxid=6239;
RN (1)
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RA MEDLINE=99069613; PubMed=9851916;
RX none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998) .
RL EMBL; 275714; CAB00060.1; -.
DR HSSP; P00962; IGR.
DR WormRep; ZC434.5; CE06580.
DR InterPro; IPR001589; Actbind_actinin.
DR InterPro; IPR004526; Glx_arch.
DR InterPro; IPR000924; Glu_cRNA-synt_1c.
DR InterPro; IPR001412; cRNA-synt_1.
DR InterPro; IPR000738; WHEP-TRS.
DR Pfam; PF00749; cRNA-synt_1c; 1.
DR Pfam; PF03950; cRNA-synt_1c_C; 1.
DR Pfam; PF00458; WHEP-TRS; 6.
DR PRINTS; PRO0987; TRNASYNTHGU.
DR TIGRFAMs; TIGR00463; glx_arch; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
SQ SEQUENCE 1149 AA; 125197 MW; CC22744C05C0A433 CRC64;

Query Match 35.9%; Score 1339; DB 5; Length 1149;
Best Local Similarity 39.5%; Pred. No. 4,1e-87;
Matches 292; Conservative 132; Mismatches 243; Indels 72; Gaps 133

QY 2 EAALSFKSDSPISIIICAAKLVLGPTTINSLAAGSAPTLQFA-----SGESLIGV 52
DB 6 ELVLKANRQOPFYASILALAAAGFSL-----EKSVQFSEKQGLALNLDOELLSD 55
QY 53 NPILLYIANGASIA-SLSCGNIDIEFGHVVEWLEYAP-----TPISGSEFENACLFVD 103
DB 56 VEIATIIAOSTPAADLDLGSIIIDFAVVDLTNFIADAVTKNDYSLLIGKDFP----- 108

```

```

QY 104 GFLASRTLVGHGLTIADIAMVSNLAGIQRRRESIRKSKKYQNLVRMFMSIDSEYKEALN 163
Db 109 -----TYLJNLSJLVADFAIFJS-----VAHNPNQJ-----AKFSQITDKVLEPTLL 150
QY 164 EVVAFAVGRGIGKSP---ASFJKEVHDSKQDPAPEVDLPBAKYGVKVCRAPEBSGYL 220
Db 151 AAANHFVULYSSAAPFTATSGRKK-----KDBEKFJELPBAEKGKVVVRPRPEASGYL 206
QY 221 HIGAKAALLNKYFAERYQRLIYFPDDTNPESKESNEFEYENLLKQIETLGICYDAVTYS 280
Db 207 HIGAKAALLNQYQOAFEGQILMFDDTNPAKENAFEHVJKEBLSMNTIYDREWTSS 266
QY 281 DYFPLTMMASJLIQOGKAYIDDPTEKQMRKERMJDSRCRNNTVYENULSLMKEMVNGT 340
Db 267 DHFEMJLTMCEKJLLEBGAFFVDDTDTETMRNEREQRODSRNRSNTPEKJLQMBEKKGS 326
QY 341 ERGMOCCVRGKJLMDQDPKSLRDPVYRXCNTDPHHRVGSKKYKVPPTYDEAFCEPVALDEG 400
Db 327 PKGLTCCVRMKJLDMKSNNGAMRDPITYRCKPEHVRGLKVKVPTPYDETCPIVDSVEEG 386
QY 401 THALRSSHYDRMAQYRIRLIDMDGRLRVEIYFSLNMYTLLISKRKJLMPQNKKEVEM 460
Db 387 THALRTTEYHORDDQYFICDALGRRRHIMEXARLINTNTYMSKKJLTFVDEGHVEGA 446
QY 461 TDPRFPYQGIYVRRLKATREALIOFLQOGASKNJLMEMWDLMTINKIILIDVCAHTAV 520
Db 447 DDPRLPYRGVWRRLIYEBGLKQFIVAOGSGSSVMMEMDKITMAFVKYIDPAARYTAL 506
QY 521 LKQORVI-FTLTNGEBEPFVRILPRHKKRBEGAKATTPFANRIWLDYADAAALNKEEYAT 579
Db 507 DSTSPJLVIEJLTDSDTDSNV-SLHPKNAEJSGDVHKGKJLLEQVDAALKESEIYT 565
QY 580 LMDWGNALVKEIKVSGSVITELVGBLHJEGSVKTKRLKJLTYLADLE---SLVPSLVEF 635
Db 566 FVNMWNKIKIGKEKKGAVITKISATLQJLNDJNTYKTKTKTYLADGVAKAEGKTIIPVLTADY 625
QY 636 DYLISKKJLEEDBFLDMJNPCSTRERIPALGDANRNIKGEIIOLEKGYRCDAPFR 695
Db 626 DHIISKALIGDEBWKQFJNFDPSVHYTKVGVBPALKVYKGGDIIOQRKGFYIVDQPYNB 685
QY 696 SSK-----PVVLPAPDG 708
Db 686 KSELSGVETPLLLAIIPDG 704

```

```

RESULT 9
Q8NNAJ6
ID Q8NNAJ6 PRELIMINARY; PRT; 717 AA.
AC Q8NNAJ6;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ35251.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ono Y.,
RA Houcha T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hiro Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagaetsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isegai T.;
RT "NEBO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK092570; BAC03916.1; -.
DR InterPro; IPR004526; GILX arch.

```

DR InterPro; IPR000924; Glu_cRNA-synt_1c.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c; 1.
DR PRINTS; PR00987; tRNA-synt_1c; 1.
DR TIGRFAMs; TIGR00463; glx arch; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Hypothetical protein.
SQ SEQUENCE 717 AA; 81797 MW; D78B40974C8D4F79 CRC64;

Query Match 34.5%; Score 1287.5; DB 4; Length 717;
Best Local Similarity 41.2%; Pred. No. 9.9e-84;
Matches 280; Conservative 127; Mismatches 244; Indels 29; Gaps 13;

QY 46 GE-SLHGNNPIILYIARAGSISLSKNDIEGHVEMLEVPATLSGSEFENACLPVDG 104
DB 43 GETTGSDALCDYITSCADQSLGGDDAE--TVKEWHIASTTKENVLEQ-TEKLD 99
QY 105 FLASRTFLVGHGLTIADIAVMSNLGIGQWESLRSKKYQNLVWPNISDEYKEALNE 164
DB 100 YIATRTYIVRVNFTIADISMAAI-----KLEVEIPAKCTFASRWYTTISA--IPSVQ 152
QY 165 VVAALFVGRGIGKSPASLSKEKVH-----DSKDPASAEVDLPGAKVGVCFAPBP 216
DB 153 TTRGVGSLRMMAKQASAPKKEIATKKGEISLDGPESWPE--LPFAEMGVVTRFPPEA 210
QY 217 SGYLIHGAKAALLNKYFAERYQGLIYRPDPTNSKESNEFEVULKDIEITGIKIDAV 276
DB 211 SGYMHIGHKAMLYYAKRYQGLILRFDDTNSKEKEPEFTSIIEDLAKIGKADLF 270
QY 277 TYSIDYFPELMEMASLIKQGAVIDDTPKEQMRKERMDGIESRCRNNTVEENLSLAKEM 336
DB 271 SHTSYFVYIADYAAQMIREGIAPMDNTDQETMKERKERKESKLRNTSPENLIPFL 330
QY 337 VNGTERGQCCVRGLDMQDPNKSIRDPIYVRCNTDPHHRVSKYKVYPTDYFACPYDA 396
DB 331 CRGEPEVDYCLRAKIDMSDNGTLRDPVLYRVFPLTHRTGDKYKAVPCYDLACPIDS 390
QY 397 LEGYTHARSSHYHRNAQYRIILODMGRVYEIYFSLNMYTLLSKRLMIVQNK 456
DB 391 IEGVTHARTETEKORDROYMWIIOALRIIRPVHLEFALNEQYTLMSRKLTVMVEHNE 450
QY 457 VEDWTPRPPTVOGIVRGKVEALIOFLQOGASKNLNMEMDLMTINKKIIDPVCA 516
DB 451 VDGMDDPRPTVKGVRKGVLEALRAFLTSQSFSEKRVVTEMDFWSNNKILERNALR 510
QY 517 HTAVLKDORVIFTLTNGSEEPFVRLIPRHKKEGAGKATTPANRIWLDYADAALNKG 576
DB 511 FMGVKWSFVTLTLTN-VQPGFIKY-PNHPKDESGVSDIAGPQVLMLERTDAQLKGE 568
QY 577 EYTLMDGNAIVKEIKVES-GVITELVGBLHESGVTKTKITWLADIIEELVPLSVEF 635
DB 569 EFTLMRWGSAKATJALRADGAVTGTGEFVNPNGDFKPT-AKVNVVAACEBNEHETVEY 627
QY 636 DYLSKKLEED--EDFLDNLPCTRREIPALGDANNRIKGGIIQLEKRGYRCAP 692
DB 638 DDLTKTKLEEDNEDFLTSKDHPTKMEYTVIVNACKGLKENQIVOLERRGFYRCVP 687
QY 693 FIRSKPVVLFALPDGROQA 712
DB 688 DTENS-PAKLIILIPDGKKA 706

RESULT 10

Q8IDK7 PRELIMINARY; PRT; 863 AA.
AC Q8IDK7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Glutamate--cRNA ligase (EC 6.1.1.17).
GN PF13_0257.

OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN (1)
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA Berriman M., Pain A., Hall N., Actin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL844509; CAD52614.1; --
KW Ligase.
SQ SEQUENCE 863 AA; 101417 MW; AEC2BD5FA7A292B8 CRC64;

Query Match 34.2%; Score 1276; DB 5; Length 863;
Best Local Similarity 39.6%; Pred. No. 8.8e-83;
Matches 273; Conservative 128; Mismatches 222; Indels 60; Gaps 10;

QY 67 SLSGKNDIEFGHV-----EWLEVPATLSGSEFENACLPVDGFLASRTFLVGHGL 117
DB 144 NISEKIDPKNNKVVNIYVQTOYEEWLDFFRSRVNKKDIFICEHLNKLHLNTFVSEYL 203
QY 118 TIADIAVMSNLAGI--GQWESLRSKKYQNLVWPNISDS-----EYKEALNEVVA 168
DB 204 TISDIFIFQYKYPHVSSTQYAKSKNYKQINRWHKILDISLYPEDAEWKNKILCLIC- 262
QY 169 FVGRGIGKSPASLSKEKVHSDK-----PSAPEVD-----LPGAKVG 206
DB 263 -----DNKANPDRKDNLKNNKSNMNTDVKCKKTVTTSYSGKLENAVIG 309
QY 207 KVCVAFAPSPGYLIHGAKAALLNKYFAERYQGLIYRPDPTNSKESNEFEVULKD 266
DB 310 NVVTRFPPEPGYLIHGAKAAPLNYYAQMVEGMLRFDDTNSKEKEPEFTSIIEDL 369
QY 267 ETLGKIDAVYITSYFPELMEMASLIKQGAVIDDTPKEQMRKERMDGIESRCRNNTV 326
DB 370 ENLGKYEKISYSSHPFLLEKCYCIDMIKMKAVADDTGVEDMRQEGIESINRNSI 429
QY 377 EENSLAKEMVNGTERGQCCVRGLDMQDPNKSIRDPIYVRCNTDPHHRVSKYKVY 385
DB 430 EKMLEFEMERKGTIEGKNCIPRAKINQSKCRDPVMTCTIVDPHHKIQFKCY 489
QY 386 TYDFACPVDALLEGYTHARSSHYHRNAQYRIILODMGRVYEIYFSLNMYTLLSK 445
DB 490 TYDFACPDIIDLEGYTHARSSHYHRNAQYRIILODMGRVYEIYFSLNMYTLLSK 549
QY 446 RKLIMFVONKVEWTDPRPTVOGIVRGKVEALIOFLQOGASKNLNMEMDLMT 505
DB 550 RKLKMFVENNVVDSVDRPPTIKGILRGLTKEALFQFILEQGPSKAGNLMQMDKWSI 609
QY 506 NKIIIDPVCARHTAVLKQRIYFTLTNGSEEPFVRLIPRHKKEGAGKATTPANRIWLD 565
DB 610 NKIIIDPPIIPYAAVANKSSILLITLTDYIQERDLHMKNSLGFCTNNYNNKYIIE 669
QY 566 YADAAAIKGEVTLMDGNAIVKEIKVESGVITELVGBLHESGVTKTKITWLADI- 624
DB 670 LEDAQTLLENSEITLIKGNIIIKIEKENGKIQINALSNFHDGFKTTKKIHLPLP 729
QY 625 EELVPLSVEFDYLSKKLEED--EDFLDNLPCTRREIPALGDANNRIKGGIIQLE 682
DB 720 QOLITCTLYEYDHLITVDKFNENDNDWTNFINFNSKETIYVVAEPPSSLSKVDSPQFE 789
QY 683 RKGYYRCAPPIRSSKPVVLFALPDGROQ 711
DB 790 RRGYFILDK-IDPHHILHLIKIPDGSK 816

RESULT 11

Q8SSE4 PRELIMINARY; PRT; 642 AA.
AC Q8SSE4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Glutamyl tRNA synthetase.
GN EC02.1210.
OS Eucephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Eucephalitozoon.
ON NCBI_Taxid=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA MEDLINE=21576510; PubMed=11719806;
RA Kacina M.D., Duprat S., Cornillio E., Metenier G., Thomarat P.,
RA Prensler G., Barbe V., Peyretallade E., Brothier P., Wincker P.,
RA Delbac F., El Aloui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Eucephalitozoon cuniculi.";
RL Nature 414:450-453(2001).
DR EMBL; AL590442; CAD25150.1; -
DR InterPro; IPR004526; GLX arch.
DR InterPro; IPR000924; Glu tRNA-synt_1c.
DR InterPro; IPR001412; tRNA-synt_1c.
DR Pfam; PF00749; tRNA-synt_1c.1.
DR Pfam; PF03950; tRNA-synt_1c.C; 1.
DR TIGRFAMs; TIGR00463; glx arch; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
SQ SEQUENCE 642 AA; 73963 MW; B2BC6D4B2C329BB2 CRC64;

Query Match 30.5%; Score 1139; DB 5; Length 642;
Best Local Similarity 44.3%; Pred. No. 3.9e-73;
Matches 235; Conservative 92; Mismatches 184; Indels 20; Gaps 8;

QY 177 KSPAPLKE-KVHSDKDPAPVVDLPGAKVKVCVRPAPEPSGYLHGHAKALLNKYPA 235
DB 120 KANRKLLEFNAGSKKEQNLIEGPSE--NVVTRPPEPGRHGHAKALNLNWFYA 176
QY 236 ERYQRLIVRPDDTNPSEKSENEFVNLKDIETLGIKDAYATVTSDFPKLMEASLIK 295
DB 177 SKGNRLVLRPDDTNPKEKEERFERGILSDLSLGINETLSTSDYFDKIIDLVPLIG 236
QY 296 QGKAYIDTPEKQAKERMKGIESRCRNVTVEENSLMKEMVNGTERGMOCCVVRGLDMQ 355
DB 237 ESKAYADNTPQEVMDERGRGVESRCSDVSESKRIFEMARGNASGY--CLRAKIDMS 294
QY 356 DPNKSLRDPVYRCNTDPPHRYGSKYKVPYTPDAPCPVDALGEGYTHALRSEYHDRNAQ 415
DB 295 SSNKRMDPVLFRVNEPPIHRTGDKRYKPYTPDFACPVIDSLGILSLRANERYDRNQ 354
QY 416 YRILQDMGLR-RVEIYEFSLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVAR 474
DB 355 YWFLDNLRNRPPIHPSRLNPFNTVLSKRLKYVDNGVSGMDPRLTLINGIKRL 414
QY 475 GLKVEALLQFIQOGASKNLIMEDKLTINKKIIDPYCAHTVAVLKQRIYFTLNGP 534
DB 415 GMMNEALREYILMOQVSOQKTCISMDKWAIRKKIDIPVSARYFCQOGDAVEASIDWTS 474
QY 535 EEPFRILPRHKKFEGAGKATTFANRIMLDVADAANKGEEVLLMDGNIVKEIYE 594
DB 475 E--YMDVPKHKKNDLGTKEVYSSQILLQEDGRVLQDNEEFTLMNGNIVASKTYE 532
QY 595 SGVITELVGLHLESGSVTKTLKLTWLDIEELVPLSLVEFDYLLSKKLEDEDFLNL 654
DB 533 NCTVTKMEVSLNPDGFKLTNKGMSVSK-RSSVYVELLAIEYGNLND---EDTEDLALRF 588
QY 655 NPCTREIIPALGDANNRINIKGEIIOLEKGYRCDA-----PPIRSSK 698
DB 589 NNSVYKEWYAESAIINVRGEVILQFERNGFYCDGFLVNLILPFTQKR 639

QY 199 DLPQAKVKVCVRPAPEPSGYLHGHAKALLNKYFARVQRLIVRPDDNPSEKSENEF 258
DB 19 DLASGKTTVTRPPEPENGYLHGHAKSLINFGIADYQGCNLRPDDNPVEDIEY 78
QY 259 VENLKDIETLGIKIDA-VTTSDFPKLMEASLIKQKAYIDTPEKQAKERK----- 313
DB 79 VDSINDVEMWGFHMSGNVRRSSDYFDQHLAVYELINKGLAYDELTPQORERGTLT 138
QY 314 MDGIESRCRNVTVEENSLMKEM-VNGTERGMOCCVVRGLDMQDNKSLRDPVYRCNTD 372
DB 139 QPKNSPYRDRSVENLALFEKMRITGPEEG-KCLRAKIDMASPFIWRDPVLIRKFA 197
QY 373 PPHRVSQKRYKPYTPDAPCPVDALGEGYTHALRSEYHDRNAQYRIIQLQDMGLR-RVELY 431
DB 198 EHHQGNKMCYIPWDFHICISDALGEGYTHSLCTLEFQDNRLVDWVLDNITIPVHPRQY 257
QY 432 EFSRLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVARGLKVEALLQFIQOGAS 491
DB 258 EFSRLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVARGLKVEALLQFIQOGAS 491
QY 492 KNLNLMEDKLTINKKIIDPYCAHTVAVLKQRIYFTLNGPEEPVRIIPRHKKEGA 551
DB 318 KQDNITENASLESICREDLNEMAPRAVVIDPVKLINENYQOEGE--NVTMHNHNDKEM 375
QY 552 GKATTFANRIMLDVAD-AAALNK-----GEVTLMDMGN-IVKIKVSSVITELV 602
DB 376 GSRQVPFGEIWIADPREEANKQYKRLVIGKEVRL--NAVYIKAEKREKDA----- 427
QY 603 GELHLEGSVKT-----TKLK--ITWLADIIEELVPLSLVEFDYLLSKK 643
DB 428 -----EGVITTFCTYADDTLSKDPADGRKTKGVHWSAHAL-PVEIRLYDLRFSVPN 481
QY 644 LEEDEDFLNDNPTCTREIIPALGDANNRINIKGEIIOLEKGYRCDAFPRISSKPV 700
DB 482 PGAADDFLSVINPESLVIKQGAPEPSLKDVAAGKAFQEREGYFCLDSRHSTAEKPV 538

RESULT 12

QY 199 DLPQAKVKVCVRPAPEPSGYLHGHAKALLNKYFARVQRLIVRPDDNPSEKSENEF 258
DB 19 DLASGKTTVTRPPEPENGYLHGHAKSLINFGIADYQGCNLRPDDNPVEDIEY 78
QY 259 VENLKDIETLGIKIDA-VTTSDFPKLMEASLIKQKAYIDTPEKQAKERK----- 313
DB 79 VDSINDVEMWGFHMSGNVRRSSDYFDQHLAVYELINKGLAYDELTPQORERGTLT 138
QY 314 MDGIESRCRNVTVEENSLMKEM-VNGTERGMOCCVVRGLDMQDNKSLRDPVYRCNTD 372
DB 139 QPKNSPYRDRSVENLALFEKMRITGPEEG-KCLRAKIDMASPFIWRDPVLIRKFA 197
QY 373 PPHRVSQKRYKPYTPDAPCPVDALGEGYTHALRSEYHDRNAQYRIIQLQDMGLR-RVELY 431
DB 198 EHHQGNKMCYIPWDFHICISDALGEGYTHSLCTLEFQDNRLVDWVLDNITIPVHPRQY 257
QY 432 EFSRLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVARGLKVEALLQFIQOGAS 491
DB 258 EFSRLNMYTLLSKRKLMPYONKKEVEDWTPRPFTVGIVARGLKVEALLQFIQOGAS 491
QY 492 KNLNLMEDKLTINKKIIDPYCAHTVAVLKQRIYFTLNGPEEPVRIIPRHKKEGA 551
DB 318 KQDNITENASLESICREDLNEMAPRAVVIDPVKLINENYQOEGE--NVTMHNHNDKEM 375
QY 552 GKATTFANRIMLDVAD-AAALNK-----GEVTLMDMGN-IVKIKVSSVITELV 602
DB 376 GSRQVPFGEIWIADPREEANKQYKRLVIGKEVRL--NAVYIKAEKREKDA----- 427
QY 603 GELHLEGSVKT-----TKLK--ITWLADIIEELVPLSLVEFDYLLSKK 643
DB 428 -----EGVITTFCTYADDTLSKDPADGRKTKGVHWSAHAL-PVEIRLYDLRFSVPN 481
QY 644 LEEDEDFLNDNPTCTREIIPALGDANNRINIKGEIIOLEKGYRCDAFPRISSKPV 700
DB 482 PGAADDFLSVINPESLVIKQGAPEPSLKDVAAGKAFQEREGYFCLDSRHSTAEKPV 538

	DP	01-MAR-2003 (TReMBLrel .23, last sequence update)
	DT	01-MAR-2003 (TReMBLrel .23, last annotation update)
	DE	Glutaminyl-tRNA synthetase, putative (EC 6.1.1.18).
	GN	Pf13_0170.
	OS	Plasmodium falciiparum (isolate 3D7).
	OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
	OX	NCB1_TaxID=36929;
	RN	[1]
	RP	SEQUENCE FROM N.A.
	RA	Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
	RA	Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
	RA	Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,
	RL	Submitted (SRR-2002) to the EMBL/genbank/DDbj databases.
	DR	EMBL, AL844509; CAD52457.1.
	KM	Aminoacyl-tRNA synthetase; Ligase.
	SQ	SEQUENCE 918 AA; 108531 MW; 1CBABA4689C80ABP CRC64;
	Query Match	18.4%; Score 685.5; DB 5; Length 918;
	Best Local Similarity	27.9%; Pred. No. 2.5e+40;
	Matches 199;	Conservative 108; Mismatches 244; Indels 163; Gaps 19
Qy	121	DIAVSNLAGIQRWES-----LRKSK-----KYONLVRFNSIDSEYKEA 161
Dd	191	DMTWKKIFFGQKKRTEPEFLDFLLYLKKENIHFFEEETRYINLHNMF----- 240
Qy	162	LNEVVAAFVGKRGIKSPASLKENVH---DSKPSAD-----EVDLPKAY 205
Dd	241	-----LNKNIVISNNKEDNTFSACNTNFIOGIIIEEDIKNNKH 279
Qy	206	GKVCFAPAEPSGYHGHAKAAALNKYFAEVQORLIYFPDTPSPKSNEFVNILKD 265
Dd	280	THVTTRPEEPGVHLGLHANSICNFGLSNKGGRTHIRFDPTNPVEIRIESIED 339
Qy	266	IETIGIKY-DAVTYSDVFPPKLMEAESLIKQAKAYIDTPREOKRKERMD----GIESR 320
Dd	340	VKMIGYDWKENHYFASNYPEOQLWEAKLIKGDADVDDQSLEELRKRNGLKDGVDSP 399
Qy	321	CRRNTVEENLSWKEMVNGTERGMQCCVRGKIDMODPNKSLRDVYYRRCNTDPHHRVGSK 380
Dd	400	YRNRIVEENLOLFENMKNGGLYEKEGEKVLRAKIMKSGMNMLRDPILIRMYKIHPRTYKK 459
Qy	381	YKVVPYVACPFRVALLEGVTHALSSSEYHDNAQYRILDOMGRLRVETYSFSLNMVY 440
Dd	460	WVIYPMDYVAHGOSISIEKITHSICTLEBETHRPLEYWFQEKLAHFKTRQIEPARLVITY 519
Qy	441	TLSRKALKMEFYQNKKVEDWDTPREPVOGIVRGGLKYEALLTOILOOGASFNULMEMD 500
Dd	520	MVMGRKLLTVNEKCYVDDMDPRPFTISGMARRGYSPLAIKDFCNKGIAGRENMIIPD 579
Qy	501	KLMWTINKKIIDPVCAHTAVLKDQEVIFLTN-GREEPFVR--ILPRHKFEAGACKATT 557
Dd	580	LLEFCVRBDMDKKAIRLPAILKPKAVITTNVADADVEIHNLVNASNHQMGOMGRNLX 639
Qy	558	FANRIWLDYADAAAANKG-----EETLMDWGNAVLYKEIKV--- 593
Dd	640	FEKEFIPIHDFOEIPODNFRFLAENRTVRLRYARCITCNEYIKXDQGGVCI--ELRCTYD 697
Qy	594	--ESGVITELV-----GELHLEGSVTKTLKLT--WLADIIEELVPLSVBPVL 638
Dd	698	PSSISGCLTNQKKDVNRQGGDGNKGGGCGNDKKVKATIHMLC-AKUSLOAEFRMYDL 756
Qy	639	ISKKKLEDED-----FLDN-----LNPECTR 660
Dd	757	FTKPPESENEDTELQKITSTAHLENNTNNNQNGKDEVASDTLLDGHLEDSESKAGMRK 816
Qy	661	EI-----PALGDANNRNKIKGGIIOLEKGYTCRAPFRRSKSPVVLPAIP 706
Dd	817	YINTNSLIHKGLVENYSTRFKIGDPIOFERVGTFPKCD-TNELPVEFNLTVP 869

AC 09LG07:2000 (Tremblrel. 15 Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DB ESTs AU097578 (E0618) .
DB Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzaeae; Oryza.
OX NCBI_taxid=4530;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (Ga3) genomic DNA, chromosome 1, PAC
RT clone: p0510F03."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AP002486; BAB03361.1; -
DR HSSP; P00962; 1GTR.
DR Gramene; Q9LG07; -
DR InterPro; IPR004514; GlnS.
DR InterPro; IPR000924; Gln crna-synt_1c.
DR InterPro; IPR001412; crna-synt_1.
DR Pfam; PF00749; crna-synt_1c; 1.
DR Pfam; PF03950; crna-synt_1c_C; 1.
DR PRINTS; PR00987; TRNASYNTHGLU.
DR TRIFAMs; TIGR00440; glnS; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
DR PROSITE; PS00178; AA TRNA LIGASE I; 1.
SQ SEQUENCE 791 AA; 89345 MW; F42BB8179B6BA0C CRC64;

Query Match 18.0%; Score 673.5; DB 10; Length 791;
Best Local Similarity 27.2%; Pred. No. 1.5e-39;
Matches 211; Conservative 139; Mismatches 300; Indels 125; Gaps 20

2 EAALSFKSDSPPI-SIIICAALVGLPLIINSLAAGSAPTLQFAGESLHGVPIL-YI 59
30 ENALVNSKVITNLAVINAEAGISGCDKTGVNLLTVAT--KIPNALVH-RPVLDIYI 84
60 ARGASIASLKGNDIEFGVVEWL-EVAPFTLSGSEFENACLIVDGFPLASRTFLVGHGLT 118
85 -----VSTKKNPAQLDAAISPLTNGPDSLDGCKFEAC-----GVGVV 122
119 IADIAVWSNL-----AGTQRMESLKKSKKYQNLVWFNSIDSEYKEA 161
123 VSIEIKSTVNEVLKHNMEALIEORYHINVGNLCSQYAKRHPHWGAKATKDEIDKKLAEI 182
162 L-----NEVNAFVGKRGIGSPAPSLKE----- 185
183 LGPTDADNVKPVKKKKKKAKVBEKKAADVTAAPSPSEELNPYSIPPOPEENFVHTETI 242
186 -----KVHDSKDPSPAEVLDLPGAKVGVCFVAPAPSGYLIHGAKKAAALLNKYFAER 237
243 FYSQGNITWRANNSKEILEKHLKATG--GKMTTRPPPPNGYLIHGAKKAMTIDGLAKE 299
238 YQGRILVRFDDTNSKESNEFVENLKDIEITLGIRYDAVYTSDFPYKLMEASLIRKQ 297
300 RNGHCYLRFDPTNPEAEKKEYIDHIOELVHMWGMEPYKVTYTSDFYALYEHAVALIRKG 359
238 KAYIDDPFKQEKREKRDGIESRCRNNTVEENLSLMKEMVANGTEGMQCCVRGLDMQDP 357
360 LAYVDHQTAEIKKEYREKKNNSPWRDRPESLKLFEIMRGLLAEGAATYLMKQDMQND 419
358 NKSJRDPPVYVCNTPDPHHRVSGSKVVPYTYFACFPVDALEGVTHALSSSEYHDNAQY 417
420 NKMSMDLIAYRIKTRPHPHAGDKKCIYSYIAHOMVDSLENITHSLCTLEBDIRPSTY 479
418 RIIDDMGLRVEIYEFSLNMYTLLSKRLMLFVQNKREVDWMDPREPTVOGIVRGILK 477
480 WLIVAGLGYCYGVMEYSRLNISNTVMSRKRLNRLVTEKVMQDMQDPRLITLAGLRAGVS 539
478 VEAIIQPILOQASAKNTLNMEDKLTMTINKKIIDVCARHFAVLKQGVIFPLTNGPEEP 537
540 STANSFPCGIGIR--SLIRVDLEVHIREELNKTASRAAVVNLPLKV--ITNLDEBK 595

QY 538 FV---RLPRKKKEGACKATTANRIWLDYADAAALNKGEVTLMDMGNAI----- 587
 Db 596 VIDDGKMPADPADASSYKVPSPRIYIEKTPRLKSDYVGLAPGKSALLRYAFP 655
 QY 598 VKELVESGVTELVEGLHLE-GSVKTKLK--ITWLADIEBELVPLSLVE---PDLASK 641
 Db 656 IKCEIVVYVGDNDPDIIEIRAEYDPSKTKPKGVLMVWQAPAGVPLVAYEFLPKLELS 715
 QY 642 KLEEDDEFLDNLNCTRR-----EIPALGDANNRNIKRGEIIQLERKGYRCD 690
 Db 716 ENPALLEWDLGDLNNSKEVIGAYAVPSLATAVL-----GDKFQFERLGYTAVD 765

RESULT 15

Q8EG26 PRELIMINARY; PRT; 556 AA.
 ID Q8EG26; AC Q8EG26;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Glutaminyl-tRNA synthetase.
 GN GLNS OR SOI786.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 Meyer T., Tsaplin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 Debey R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
 Mueller U., Knout H., Gill J., Utterback T.R., McDonald L.A.,
 Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
 "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AE015623; AAN54839.1; --.
 DR TIGR; SOI786; --.
 KW Aminoacyl-tRNA synthetase; Complete proteome.
 SQ SEQUENCE 556 AA; 64103 MW; 3A7E689DDACCAFFFO CRC64;

Query Match 17.9%; Score 669; DB 16; Length 556;

Best Local Similarity 34.7%; Pred. No. 1.8e-39; Matches 182; Conservative 77; Mismatches 223; Indels 42; Gaps 13;

QY 199 DLPGAQKVCVAFPEPSGYLHIGAKALINKYFAERYOGLIVRPDDTNPSESNEF 258
 Db 21 DLKSGKHNVQTRFPEPENGYLHIGAKSICLNFGLARDYQGLCNLRFPDTPNPEKEDIDY 80
 QY 259 VENLIKDIETLGIKYDA-VTYTSDYFPKLMWAESLIQOKAYI---DTPKQMKRER 313
 Db 81 VNSIQADVRLWLGFGQDGEVRYSSNYFDQLHQYAVELINKGLAYVCFNLNADETREYRGTLK 140
 QY 314 MDGISRCRNNVVENLSLMKEMVNGTERGMQCCVRGLMDQDPKSLRDPVYRCNDP 373
 Db 141 EPGKSPYRDPFVEENLRFGMRLGEEFEGECALRAKIDMASPFMCMDPVIYIRIPAH 200
 QY 374 HHRVSGTKVPTYPYFACPFVDALGVTALRSSEVHDNNAQYRILODM---GLRVE 429
 Db 201 HKQTDDKCIYPMYFTHCISDLAHHITSLCTLEFQDNRLYDWLNDLDDFOAPNRTR 260
 QY 430 IYEFRLMMVYTLTKSKRLMFWQNKKYEDWTDPRFPPTVQGIIVRGKLYEALIQFILOQ 489
 Db 261 QYEFRLMLVEYTLMSKRKLNDLVTRKLVGWDDPRMPTIAGLRRGYTPASIRFCQRIQ 320
 QY 490 ASKNIATLMEWDLMTINKKIIDPVCARHTAVLKQQRVIFTLTNGPEEPFVRI-LPRKKKF 548

Db 321 ITKQENMIAMLDACIRBELNEHAPRAMAVIRPLKV--IENYPEGQIETIQASHPSD 378
 QY 549 EGACKATTANRIWLDYAD-AAALNK-----GEEVTLMDMGNAIYKEIKVE----- 594
 Db 379 ESMGTRELAFGRLEIFIDVADFREEBANKQYKELVNGKEVRL---NAVY-ITAEKCDKS 433
 QY 595 SGVIT-----ELVGEHLGSGVTKTKITWLADIEBELVPLSLVEFDYLISSKLEE 646
 Db 434 EGNITTYCSYDADDTLGNPADG--KRYKGVITHV-EATTKAPQAFRIYQRLFTDPDPA 490
 QY 647 DEDFLDNLNCTRRREIPALGDANNRNIKRGEIIQLERKGYRCD 690
 Db 491 AETVDEVLPNSLEVNGLVASLANAPAEKAYQFEEREGYFCAD 534

Search completed: January 25, 2004, 17:13:14
 Job time : 198 secs